2000

Gas processes in galaxy formation

Mirjam Jonkman
University of Wollongong

GAS PROCESSES IN GALAXY FORMATION

A thesis submitted in fulfilment of the

requirements for the reward of the degree

MASTER OF SCIENCE (HONOURS)

from

THE UNIVERSITY OF WOLLONGONG

by

Mirjam Jonkman

Department of Physics

March 2000

Abstract

This work consists of two parts. The first part investigates the effect of an assumption in Nulsen and Fabians semi-analytical model for galaxy formation. They assume that star formation in low mass systems is regulated by supernova explosions which shut it off by expelling the rest of the gas from the galaxy. The sensitivity of the model to their assumption that this happens when the energy released by supernovae exactly equals the binding energy of the remaining gas is tested here. It is found that the outcomes of the model are insensitive to this assumption, with the exception of the ratio of elliptical to spiral galaxies, which is also sensitive to a number of other model parameters. It is concluded that their assumption has a relatively minor influence on the results of their model.

The second part of this work involves the the in-shock cooling problem. Lack of resolution causes numerical shocks to be much thicker than real shocks, resulting in excessive cooling in simulated shocks. The problem was investigated using 2- and 3-dimensional SPH models for the formation of a disc galaxy. In-shock cooling was demonstrated, based on the strong correlation between viscous heating and radiative cooling rates. Several approaches are considered to alleviate the problem. It is found that reducing the radiative cooling by the amount of viscous heating does not produce satisfactory results. This is the result of the rapid rise of the cooling rate with temperature for low temperatures. The second approach, turning cooling off in shocks, requires a criterion to decide whether a particle is shocking or not, but no universally applicable criterion was found. All dimensionless criteria tested fail to distinguish between shocks and homologous collapse. In-shock cooling can significantly be reduced by switching cooling off for particles which have heating rates above a certain value, but this criterion is highly dependent on the system being simulated and is not useful in flows involving a wide dynamic range of scales. Prospects of a general solution to this problem look poor.

Acknowledgements

I would like to sincerely thank all of the people who helped me during the research and writing of this thesis. First I am deeply grateful to my supervisor, Dr. Paul Nulsen, for his encouragement in times when that was really needed, his consideration and for useful discussions and the occasional good argument. I would like to thank Peter Maguire for making his axisymmetric code available and for his emails explaining SPH. Thanks also to Peter Ihnat and Peter Anthony for keeping the computer going. The staff and students of the department deserve thanks for making me feel welcome and for sticking together in difficult times. A special thanks goes to Vince McIntyre for his never ending interest in the project, his enthusiasm for astronomy and for always offering the observers perspective. I am also grateful to Prof. Rian Dippenaar, for giving me every opportunity to finish this thesis while being his PhD student. Finally, and most importantly, I would like to thank Friso for his support throughout and especially during the last few months.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

Galaxy formation is the subject of intense study, both by observers and by theoreticians. The process of galaxy formation is complex and theoretical studies therefore rely primarily on numerical simulations. In order to evaluate models of galaxy formation, the outcomes of simulations have to be compared with observed properties of galaxies. In principle, the models should account for all properties of galaxies, but, in practice current models attempt to account for the main properties, like morphology, luminosity and the gross dynamical properties given by the Faber-Jackson and Tully-Fisher relations. These properties are discussed briefly here.

The visible matter in the universe consists of stars, gas, dust and active galactic nuclei. Stars and active galactic nuclei shine in their own right, gas and dust are mainly made luminous by other objects. Most of the luminous matter in the universe is in the form of stars and most stars occur in galaxies.

Galaxies vary widely in their properties and can be classified accordingly. The standard classification [15] [16] divides galaxies according to their morphology. Two main categories are distinguished, elliptical galaxies and spiral or disc galaxies. Elliptical galaxies are smooth, featureless distributions of stars. Spiral or disc galaxies, the type of our own galaxy, have a prominent disk with a central bulge and often contain spiral arms.

Another distinction that can be made is according to size. Galaxies are divided into two categories by size, dwarf galaxies and normal galaxies. Dwarf galaxies differ from normal galaxies not just in their size but also in their morphology and are thought to have formed in a different manner [38]. Masses of galaxies

vary from $10^6 M_\odot$ for small dwarf galaxies to more then $10^{13} M_\odot$ for giant elliptical galaxies.

Galaxies do not normally occur in isolation. Most of them belong to groups or clusters [5]. Clusters are the largest virialized structures in the universe and have typical sizes of a few megaparsecs. In addition to the galaxies, clusters also contain a large amount of hot intracluster gas, with temperatures typically in the range of 10^7 to $10^8 K$ [47]. The proportions of the different types of galaxies depend on the environment. In low density regions the majority of galaxies are disc galaxies and only a small fraction of all galaxies are ellipticals, while in dense clusters of galaxies the number of spiral galaxies is smaller and the fraction of elliptical galaxies rises to 40% [108].

The galaxy luminosity function, ϕ, is defined so that $\phi(L)dL$ is the number of galaxies per unit volume with luminosities between L and $L + dL$. This is one of the main tests for galaxy formation models. The luminosity function differs for the different types of galaxies and therefore also varies with the density of the environment [80] [14]. A commonly used functional form for the luminosity function of elliptical galaxies is the Schechter function:

$$\phi(L) = \frac{n_*}{L_*} \left(\frac{L}{L_*}\right)^{\alpha} \exp\left(-\frac{L}{L_*}\right)$$

where n_*, α and L_* are parameters. Observed values are, approximately, $n_* = 1.2 \times 10^{-2} h^3\,\mathrm{Mpc}^{-3}$ where h is the value of the Hubble constant in units of $100\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\alpha = -1.25$ and $L_* = 1.0 \times 10^{10} h^{-2} L_\odot$ [16].

Other properties of galaxies and clusters which can be used to test galaxy formation models are mass and temperature functions. The mass function gives the number of bound objects per unit volume in the mass range $(M, M + dM)$

[53] [63] [135]. Mass functions are less accessible to observation than luminosity functions due to the difficulties of determining masses. For groups and clusters, temperature functions give the number of systems per unit volume in which the hot gas lies in the temperature range $(T, T + dT)$ [63].

The gross dynamical properties of galaxies are described as the relationship between luminosity and velocity dispersion or rotational velocity. They are given by the Faber-Jackson relation $v_c = 220(L/L_*)^{0.25}$ km s^{-1} for elliptical galaxies and the Tully-Fisher relation $v_c = 220(L/L_*)^{0.22}$ km s^{-1} for spiral galaxies, where v_c is the circular velocity.

The luminous matter, discussed so far, is not the only form of matter. Estimates of cluster mass based on their dynamics [39] [40], the distribution of the hot gas (e.g. [58] [36]), the Sunyaev-Zel'dovich effect (e.g. [56]) and gravitational lensing [116], all agree that the total mass greatly exceeds the mass of the luminous components. It is estimated that considerably more than half of the matter in clusters is dark. There is also a variety of evidence for dark matter in galaxies [134] [16].

The nature of this dark matter is unclear, although a great many candidates have been proposed, ranging from brown dwarfs and low mass stars to exotic elementary particles. They are often divided into two categories: MACHO's (Massive Compact Halo Objects), which are assumed to be baryonic and WIMPs (Weakly Interacting Massive Particles), which are assumed to be non-baryonic. It is now believed that there is both baryonic and non-baryonic dark matter [25]. There is evidence for the existence of MACHO's 17 candidate microlensing events were found for objects in the direction of the Large Magelanic Cloud [3].

Estimates for the MACHO mass depend on the model, but lie between $0.15M_\odot$ and $0.9M_\odot$. It is estimated that the total halo mass in MACHO's is about $9.0 \times 10^{10}M_\odot$, which is about 7 % of the total mass of the Galaxy [74].

The possible forms of non-baryonic dark matter can be further subdivided into neutrino's, which are known to exist although their mass is uncertain, and exotic elementary particles, which have not been observed yet. The relevance of these different forms of dark matter will be further discussed in section 2.

1.2 Galaxy Formation

1.2.1 Cosmology

Cosmology is the study of the history of the universe as a whole. The standard view of cosmology can be summarized in four points [106]. First, the universe is homogeneous and isotropic on large scales. This means that we do not inhabit a special location in the universe and that there is no preferred direction. Second, the universe is expanding with the rate given by the Hubble constant, H_0. Hubble demonstrated in 1929 that the recession velocity of a distant galaxy, v, is proportional to its distance, l, with $v = H_0 l$. Current determinations give $H_0 \simeq 70$ [88]. Third, the dynamics of the universe can be described by Einstein's general theory of relativity. Fourth, the universe started in a hot dense state, the big bang. The cosmic microwave background radiation is a relic of thermal radiation from the hot early stage of the big bang.

The first and third points are described mathematically by the Friedmann-Robertson-Walker models. These models leave three possibilities. If the average density of the universe, ρ, equals the critical density, ρ_c, the universe is flat and

the geometry is Euclidian. For $\rho > \rho_c$ the universe is closed and the topology can be described as the surface of a four dimensional hypersphere, and for $\rho < \rho_c$, the universe is open. The density parameter Ω is defined as

$$\Omega = \frac{\rho}{\rho_c}$$

so that the universe is open, flat or closed if $\Omega < 1$, $= 1$ or > 1 respectively. Theoretically, the model with $\Omega = 1$ is the most appealing. It is the simplest model and, as the universe expands, the value of Ω diverges from 1 if it is not equal to 1. Given that the universe is close to flatness now, Ω must have been extremely close to 1 at early times. Current observations favour a flat universe, with matter contributing $\Omega_m \simeq 0.3$ to the density parameter and the balance being due to the vacuum energy density [107] [104].

Models for galaxy formation must fit within the standard cosmology. The early Universe is thought to have consisted of nearly uniform expanding matter in which small density perturbations grew under the influence of gravity. Eventually these density perturbations decoupled from the general expansion and recollapsed. The origin of the fluctuations is unclear. One possibility is that they were formed by inflation [90], a phase transition in the early universe during which exponential expansion occurs. Inflation predicts a scale invariant Gaussian spectrum of density fluctuations.

How the density fluctuations evolve depends on the nature of the matter in which they are formed. The constraints from primordial nucleosynthesis limit the contribution of baryons to the mass density. Based on primordial deuterium abundances, Burles and Tytler [22] [23] [73] find $\Omega_b h^2 \simeq 0.019$ where h is the value of the Hubble constant in units of $100\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. This means that

most matter in the universe is non-baryonic dark matter.

Non-baryonic dark matter can be divided into two categories, hot dark matter and cold dark matter. Hot dark matter (HDM) consists of fast moving particles. Neutrino's are a candidate for this form of dark matter if they have a non-zero rest mass. The attractiveness of this form of dark matter is that neutrino's are known to exist. Streaming of hot dark matter tends to wipe out density fluctuations on the scale of galaxies and smaller. This means that galaxies form in a 'top-down' manner in hot dark matter models. The first structures to form are large sheet-like protoclusters (Zel'dovich pancakes) which break up into individual galaxies [104].

Cold dark matter (CDM) consists of pressureless exotic particles. The disadvantage of models with cold dark matter is that they assume a hypothetical form of matter which has not yet been observed. However, although CDM models don't reproduce all the observed structures perfectly, they lead to more realistic scenarios for galaxy formation than HDM models. In CDM models structure develops 'bottom-up'.

The density perturbations are usually assumed to be a random superposition of density fluctuations on different scales, so that the highest amplitude perturbations occur on the smallest scales in CDM models. The collapse time of a density perturbation depends on its amplitude, if the amplitude is higher the collapse time is shorter. This means that the smallest structures collapse first. These small structures merge into larger ones as they collapse, which in turn merge into still larger ones and so on. This process is called a hierarchical collapse.

Usually a scale invariant power spectrum is assumed for the initial density

fluctuations. This spectrum is modified by pressure and dissipative effects before matter and radiation decouple at $z \simeq 1000$. The scale invariant power spectrum has the advantage of having only one free parameter, its amplitude. The amplitude of the spectrum is conveniently specified by σ_8, the amplitude of the mass fluctuations in $8h^{-1}$Mpc spheres, where h is the value of the Hubble constant in units of $100\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. σ_8 can be normalized to the measurements of the fluctuations in the Cosmic Background Radiation on 10° angular scales by the COBE satellite [21] [55] [37]. CDM models with scale invariant fluctuations reproduce a number of important features of the Universe. Galaxies are formed by hierarchical clustering and numerical simulations produced large scale structures which are similar to what is observed. However, CDM models do not reproduce all observations [37].

1.2.2 Gas Dynamics

Gravitational collapse alone is not sufficient to describe the formation of galaxies. In a hierarchical collapse, substructure, i.e. structure formed during previous collapses, is wiped out (although this picture may be incomplete, see [50]). A pure hierarchical collapse results in a self-similar distribution of matter, where collapsed structures look the same on every scale. If gravitation was the only relevant process, there would be no galaxies in groups and clusters. There would also be no preferred size or mass. The visible parts of normal galaxies, however, have a characteristic mass of about $10^{11}\mathrm{M}_\odot$ and a characteristic size of about 10 kpc. To explain the observed sizes of galaxies and their survival through later stages of the collapse hierarchy, gas processes need to be considered.

Rees and Ostriker [110] demonstrated the importance of the competion between shock heating and radiative cooling in a collapsing gas cloud. They showed that the collapsed cloud could only be hot above a certain mass or radius. However, their model only considered a self-gravitating gas cloud and did not include dark matter. White and Rees [139] were the first to include dark matter in their model. They assumed that the dark matter has undergone hierarchical clustering in which, at each stage of the collapse, substructure was wiped out. The luminous content of galaxies, however, is much more concentrated than the dark matter. When the gas falls into potential wells provided by the dark matter it is heated but it can cool again by emitting radiation. If the gas can cool sufficiently, it contracts further and finally fragments into stars. Galaxies thus formed as concentrated luminous cores embedded in extensive dark halos. Dissipation due to radiative cooling allows the galaxies to become sufficiently concentrated to survive disruption of their halos in groups and clusters of galaxies.

Star formation and the associated feedback from supernovae is an important process in galaxy formation. If the gas cools sufficiently it can fragment and form into stars. It is not clear what determines the masses of these stars. The minimum mass of gas that can collapse is the Jeans mass, the mass at which gravitation can overcome thermal pressure. However, stellar masses bear no simple relation to the Jeans mass. A proportion of the stars turn into supernovae which release their energy into the interstellar medium. It is generally believed that feedback from supernovae plays a role in the formation of dwarf galaxies. One theory is that supernova explosions remove the remaining gas from the galaxy in winds, thereby stopping further star formation and preserving gas until later collapses

[38].

Gas processes are also relevant in the formation of disc galaxies. Protogalaxies obtain their angular momentum by tidal torques before condensing out of the general expansion [103]. This is usually quantified in terms of the dimensionless parameter $\lambda = J|E|^{1/2}G^{-1}M^{-5/2}$ where J,E and M are the total angular momentum, energy and mass of the system respectively, and G is the gravitational constant. Numerical simulations give values of $\sim$ 0.08 for λ. Given these values, Fall and Efstathiou [46] found that collapse factors of 10 or more are required to explain the rapid rotation of the matter in galactic discs. Nulsen and Fabian [99] argued that such a large collapse factor is not necessary if a disc forms from hot gas and viscosity is taken into account. The reasoning behind this is that the viscosity in the hot gas tends to cause it to corotate. Thus, when the hot gas cools and flows inward it leaves most of its angular momentum behind in the remaining gas. In this way the last gas to cool rotates very rapidly and this gas can form the disc.

1.3 Simulations of Galaxy Formation

The theory of galaxy formation is studied with numerical techniques. Broadly speaking the methods used can be divided into two categories, N-body methods and semi-analytical methods.

In an N-body model matter is represented by a set of particles which move under the influence of their mutual gravity. The particles sample the distribution of matter but cannot necessarily be identified with physical objects. Forces between these particles are calculated and the equations of motion are integrated

to simulate the way the matter moves. N-body methods can be combined with hydrodynamic computations, in order to include the gas component. One way often used to do this is Smooth Particle Hydrodynamics (SPH). SPH models the fluid elements as discrete particles, carrying gas properties like temperature and pressure. Local values for these properties can be found by taking a weighted average of the properties of the surrounding particles. The particles move under the influence of gravity and pressure forces. N-body hydrodynamic codes are treated in more detail in chapter 4.

In principle, the combination of an N-body code with hydrodynamic simulations should give the most accurate and complete results. In practice, however, there are number of limitations. The main problem is the lack of resolution. Since the total number of particles is limited by the available computing facilities, the number of particles in individual structures is often too small to model normal galaxy formation. This problem is compounded by the inclusion of gas, since gas processes often need to be resolved on a smaller scale than gravitation alone. N-body methods are slow because of the large number of calculations which have to be performed. This is a limitation which makes it difficult to incorporate processes, like star formation, that are poorly understood as it is not possible to try out very many different possibilities.

Semi-analytical methods reduce models for structure formation to a set of simple rules by using a combination of simplified physics and heuristic reasoning. They consist of two components, the merger tree and the physical collapse model. The merger tree, which provides the masses, times and history of individual collapses is constructed based on an assumed cosmology. The physical collapse

model determines the outcome of individual collapses. Based on some physically motivated rules, quantities like the amount of star formation are calculated. The results can then be used as input parameters for subsequent collapses. The final outcome is a set of predictions for the observable properties of galaxies, such as luminosity functions and the Tully-Fisher relation. The main advantage of this class of models is that they are much less computationally intensive than N-body methods. This makes it possible to do a large number of simulations in a limited time span. In this way a large parameter space can be explored. The limitation of these models is that they always involve a large number of assumptions.

1.4 This Work

This work consists of two parts. The first part involves investigating the effect of an assumption in a semi-analytical model, used by Nulsen and Fabian (NF) [100] [101], to simulate galaxy formation. NF combine the merger tree with their particular expectations of gas processes, which differ in a number of areas from those of other authors. The aspect of their model which is investigated here is their treatment of the effect of feedback from supernovae. NF assume that in low mass systems star formation is regulated by supernova explosions which eject the remaining gas from the galaxy, thereby stopping further star formation. One of the assumptions they make is that this happens when the energy released by supernova explosions exactly equals the binding energy of the remaining gas, i.e. the energy required to take the gas to infinity. Here, the influence of that assumption is tested in two different versions of the model, the model described in [100] which has a flat cosmology and the model described in [101] which has

an open cosmology.

The second part of this work involves the attempt to find a solution for the in-shock cooling problem. The in-shock cooling problem was first noted by Nulsen and Fabian [99] and later studied by Maguire [78]. The problem is caused by limited spatial resolution in numerical simulations. In reality shocks are only a few particle mean-free-path lengths thick but, due to the lack of resolution, they can be more than four orders of magnitude thicker in numerical simulations [78]. Since the speed of shock propagation is not much affected by numerical resolution, the time to travel through the shock is much longer in simulations than it is in reality. While in reality the shock crossing time is so short that cooling can be ignored, the gas has time to cool significantly during a numerical shock. The result is that in simulations of galaxy formation insufficient hot gas is produced. While the last point is well known, it is not usually attributed to the in-shock cooling problem. Here, it will be demonstrated that in-shock cooling is indeed responsible for the lack of hot gas and different methods to ameliorate the problem will be explored. This is done in two models for the formation of a disc galaxy, a 2 dimensional axisymmetric SPH model written by Maguire but revised and modified by the author, and in a 3 dimensional SPH model written by the author.

In chapter 2 semi-analytical methods and in particular the Nulsen and Fabian model will be described in more detail. Their theory of galaxy formation will be explained and the reasons for testing the model assumption will be given. Chapter 3 gives the results of relaxing the assumption for the two versions of the model. The second part of this work is described in chapters 4 to 6. In chapter 4 a review

of N-body codes and Smooth Particle Hydrodynamics is given and in chapter 5 the two and three dimensional galaxy formation codes are described. Chapter 6 discusses in-shock cooling and evaluates the results of a number of proposed methods for alleviating the problem. Finally conclusions are given in chapter 7.

Chapter 2

The Semi-analytical Model

2.1 Introduction

Semi-analytical models use a combination of simplified physics and heuristic reasoning to reduce models for structure formation to a simple set of rules. They consist of two components, the merger tree and the individual physical collapse model. The merger tree gives the masses, times and histories of the individual collapses. Most methods of constructing the merger tree are based on Press-Schechter theory [109] which in turn is based on the Gunn and Gott model for spherical infall [57]. According to this model the mass inside radius R will be incorporated into a collapsed object when the overdensity, $\delta = \rho/\overline{\rho} - 1$, determined from linear theory and smoothed with a top-hat filter of radius R, exceeds a critical threshold δ_c. Here ρ is the local density and $\overline{\rho}$ the mean density of the universe. Given an initial spectrum of density fluctuation, the number of objects per unit volume lying in the mass range $(M, M + dM)$ can thus be found for any redshift. The theory has been extended by Bond et al. [18] and Bower [19] who derived expressions for the fraction of material in objects of mass M_1 at z_1 that was part of objects of mass M_2 at z_2. Using this, a detailed merging history of dark matter halos can be found.

One simple method of constructing the merger tree is the Cole and Kaiser block model [32] [31]. Cole and Kaiser use the power spectrum of density fluctuations to compute the variance of the density fluctuations as a function of mass scale. A large block of the universe, containing mass M is given a density perturbation drawn from a Gaussian distribution with zero mean and standard deviation $\sigma(M)$. This block is then divided into halves. An extra perturbation is added to one half and subtracted from the other half. The standard deviation

of the Gaussian from which this perturbation is drawn is such that adding it in quadrature to $\sigma^2(M)$ produces a variance $\sigma^2(M/2)$. This procedure is repeated on ever smaller blocks until the original volume is divided into 2^N volumes, each of mass $M/2^N$. One limitation of this model is that the total mass of the blocks can only grow by powers of 2, introducing some sensitivity in the model to the choice of the mass steps. Another limitation is that, when a block is divided into two, one and only one half has a larger density than the original block. This means that only one of the two smaller blocks can collapse before the larger block collapses, with the result that mergers of two collapsed systems of nearly equal mass do not occur. Also the method gives no spatial information.

The various models differ widely in their handling of individual collapses [61] [71] [119] [75] [136] [31]. Because processes like star formation are poorly understood, semi-analytical models always contain a number of assumptions and parameters. Some major areas where models differ include the way the amount of star formation is determined, the effect of the feedback due to star formation and the inclusion and the effect of cooling flows. Authors also differ in the importance they attach to different outcomes of the model. Some outcomes which are mentioned are: luminosity functions, the Tully-Fisher relations, mass distribution, colors of galaxies and properties of cluster gas.

A few comparisons between semi-analytical techniques and N-body methods have been made [118][11]. Somerville et al. [118] compared an N-body model with Press-Schechter theory and extended Press-Schechter theory [109] [19]. They found that although agreement between these two methods is not as good as sometimes thought, the halo merging histories constructed with the extended

Press-Schechter formalism should provide a reasonably reliable framework for semi-analytical modelling of galaxy formation.

Comparisons including the gas processes are more difficult, due to our lack of understanding of what governs the rate of star formation and the initial mass function and the limited resolution of N-body codes. Benson et al. [11] compared a Smooth Particle Hydro-dynamics (SPH) model with two versions of their semi-analytical model, the full version and a stripped down version. In the stripped down version processes which were not included in their SPH model (like star formation and supernova feedback) were turned off and the cooling was adjusted to mimic the SPH model. They concluded that the overall agreement between the different methods was good, although some properties differed by a factor 2. The resolution of the SPH model had to be adjusted in order to obtain the right density for the cold gas.

Hybrid models, in which N-body simulations were used to select dark matter halos and semi-analytical techniques were used to model the galaxies within these dark halos, have also been applied [70] [10]. These model provide spatial information while maintaining the main advantages of semi-analytical model

Because of their versatility semi-analytical methods can be used to study a variety of processes. The Nulsen and Fabian model (hereafter NF), which is used for this work, has been applied to investigate the effect of cooling flows [100] , the effect of supernova feedback on cluster properties [141], the formation of quasars [101] and damped Lyman alpha absorption systems [98]. Here the influence of one of their assumptions is investigated in two versions of the model, the version described in [100] and the version described in [101]

2.2 The Nulsen and Fabian Model

2.2.1 The Merger Tree

NF use the Cole and Kaiser [32] block model to construct the merger tree. The mass of the smallest regions simulated in the block model is $1.5 \times 10^{10} M_{\odot}$ and there are 20 levels of collapse hierarchy. This means that the total mass in one collapse tree is $7.9 \times 10^{15} M_{\odot}$, larger than any present day galaxy cluster. The first model [100] assumes a CDM power spectrum of density fluctuations in a flat dust cosmology. H_0 was taken as $50 \mathrm{kms}^{-1}\mathrm{Mpc}^{-1}$. Since it is impossible to satisfy all the constraints simultaneously, a normalization of $\sigma_8 = 1$ was chosen, which leads to an excessive number of rich clusters of galaxies. The baryon fraction is 0.3., in accordance with gas fractions in clusters, although this is not in agreement with the limits from primordial nucleosynthesis.

The second model [101] has an open cosmology. The density parameter $\Omega = 0.3$ and the baryon fraction 0.25. The Hubble constant, H_0 is again taken as $50 \mathrm{kms}^{-1}\mathrm{Mpc}^{-1}$ and $\sigma_8 = 1$.

2.2.2 The Individual Collapses

The physical collapse model was first described in [99]. With a few minor changes this is used in both models. The collapse produces a halo of dark matter which is taken to be a perfect isothermal sphere ($\rho(R) \propto R^{-2}$). This halo is truncated at the radius R_{200}, where the mean density within the sphere is 200 times the background density in a Einstein-de Sitter universe at the time of the collapse. The gas collapses with the dark matter. During the collapse it may be shock-heated and, by emitting radiation, it can cool again. If the gas gets sufficiently

cold and dense, it can form stars.

NF calculate the amount of gas which turns into stars by considering a notional collapse in which the gas does not emit any radiation. In this collapse the gas would form a hydrostatic atmosphere. Like the dark matter the gas is assumed to form an isothermal sphere which is truncated at R_{200}. The temperature of the gas, however, may be different from the virial temperature, $T_{\rm vir} = m\sigma^2/k$, where m is the mean molecular mass, σ the line of sight velocity dispersion of the dark matter potential, and k Boltzmann's constant. This is the result of gas processes during earlier collapses, in particular the injection of energy into the gas by supernova explosions. For isothermal gas at a temperature $T_{\rm gas}$ the density follows a profile of $\rho(R) \propto R^{-2\beta}$, with $\beta = m\sigma^2/kT_{\rm gas}$. β is a dimensionless parameter which generally lies in the range 0.5 to 1. If the temperature of the gas is equal to the virial temperature β equals 1. If $\beta < 0.5$ in the model the gas is not gravitationally bound.

The heated gas loses its energy by different radiative processes, depending on the temperature of the gas and its metallicity. In all cases, however, the cooling time, i.e. the time it would take the gas to lose all its thermal energy at the current rate of cooling, is inversely proportional to the density of the gas. This means that the inner parts of the isothermal sphere, where the density is highest, cool fastest. The cooling time is $t_{\rm cool} = \frac{3}{2} n_{\rm T} kT/n_{\rm e} n_{\rm H} \Lambda$, where $n_{\rm T}$ is the total particle density, $n_{\rm e}$ and $n_{\rm H}$ are the electron and proton densities respectively, and Λ is the cooling function, which in these models is based on the results of Böhringer and Hensler [17].

NF now separate the gas into two parts, one part which, after the collapse, will

be cold enough to turn into stars, and one part which will remain hot. They do this by taking the hydrostatic atmosphere, that would be formed in their notional non-radiative collapse, and finding the radius where the cooling time equals the gravitational free fall time. The gravitational free fall time, the time it takes a particle to fall to the centre of the isothermal sphere, is $t_{\rm grav} = \frac{R}{v_{\rm c}}\sqrt{\frac{\pi}{2}}$ where $v_{\rm c}$ is the circular velocity. If the cooling time is shorter than the gravitational free fall time, any shock heating during the collapse is transient, so the gas ends up cold and is immediately able to form stars. If the cooling time is longer than the free fall time the gas forms a hot hydrostatic atmosphere. It may cool later and take part in a cooling flow. The radius at which the cooling time equals the free fall time is called $R_{\rm CF}$. Gas within $R_{\rm CF}$ is said to be catastrophically cooled, it is cold immediately after the collapse and forms into stars. Gas outside $R_{\rm CF}$ remains hot.

A fraction of the catastrophically cooled gas will turn into the massive stars, above $\sim 8{\rm M}_{\odot}$, which form type II supernovae. These supernovae release their energy into the remaining gas and enrich it with metals. NF assume that if their combined energy input is sufficient, the gas which has not yet turned into stars is ejected from the proto-galaxy and a dwarf galaxy is formed. This is different from the treatment of other authors who assume that supernova feedback regulates the rate of star formation locally. As NF pointed out in their earlier papers [99], in order to regulate star formation, supernova feedback must propagate throughout the cold gas in less than about one dynamical time. If this is the case then the amount of supernova energy injected into the gas would also be more than sufficient to unbind the rest of the gas, thereby stopping further star formation.

This is similar to the model proposed by Dekel and Silk [38] for the formation of dwarf galaxies.

In those cases where the energy from supernovae is not sufficient to expel the rest of the gas from the galaxy, all the catastrophically cooled gas turns into stars. The remaining gas forms a hydrostatic atmosphere with a density profile of $\rho(R) \propto R^{-2\beta}$, where again $\beta = m\sigma^2/kT_{\rm gas}$. The temperature of the gas is recalculated to adjust for the energy released by supernovae during the earlier part of the collapse. Although the hot gas could not cool in less than the gravitational collapse time, some of it is still able to cool before the next collapse. According to NF this gas takes part in a cooling flow and forms baryonic dark matter [99] [45] [129], which they identified with the MACHO's. Given the results from microlensing, [2] [9] [3], the masses of normal stars, this interpretation is problematic. A normal galaxy is assumed to form if the mass deposited by the cooling flow exceeds the mass of stars formed from the catastrophically cooled gas.

There is a slight difference between the two models in the way the time available to cool the hot gas is calculated. In both models spiral galaxies are assumed to form if all the hot gas can cool before the next collapse or before the present, whichever comes first. The reasoning behind this is that viscosity in the hot gas tends to cause the gas to corotate. Thus, when the hot gas cools and flows inward it leaves most of its angular momentum behind in the remaining gas. In this way the last gas to cool rotates very rapidly and this gas forms the disc. The model does not include star formation in the discs of spiral galaxies. All the hot gas is assumed to form baryonic dark matter.

If the collapse contains galaxies formed during previous collapses they are treated as follows. Dwarf galaxies are destroyed with their stars contributing to the spheroid of the resulting system. Normal galaxies, however, survive any collapse and any system with more than one normal galaxy is called a cluster. The model does not allow for mergers between normal galaxies. An elliptical galaxy can be converted into a spiral galaxy if all its hot gas cools before the next collapse or before the present. There is, however, no mechanism to convert spiral galaxies into ellipticals.

2.3 Reasons to Change the Model

An important assumption in the NF model is that star formation continues until either all the catastrophically cooled gas is turned into stars or the energy released by supernovae explosions is sufficient to eject the rest of the gas, i.e. the remaining cold gas and any hot gas around it. The energy released by supernovae is considered to be sufficient to eject the remaining gas if it exactly equals the binding energy of the gas which has not yet turned into stars. This is an idealized situation. It relies on the assumption that feedback reaches all star forming areas simultaneously and shuts of further star formation as soon as the condition is satisfied. If feedback propagates on the same time scale as star formation, however, it is quite possible that more stars are formed than precisely necessary to expel the remaining gas, since star formation may continue even after sufficient energy is released into the interstellar medium. Supernovae are not likely to be uniformely distributed and this can increase the time necessary for supernova feedback to propagate throughout the galaxy. This results in more

star formation and a larger amount of energy in the ejected gas.

Another reason why the assumption represents and overly simplified case is the complexity of the interaction of supernovae with the interstellar medium [29] [30] [138] [140]. The propagation speed of supernova feedback will depend on properties of the interstellar gas. Supernova explosions result in a shock wave and, in the dense regions behind the shock, energy can be radiated away very efficiently. This may lead to enhanced star formation, opposing the effect of supernovae feedback in ending further star formation.

There is also the possibility that the energy required to shut off star formation is less than the binding energy of the gas, i.e. the energy required to send the gas to infinity. This is because the gas may be removed from the galaxy without being expelled to infinity. Star formation could end before the total supernova energy is equal to the binding energy. This gas may fall back eventually, but not before the next collapse, or it may fall back but not be dense enough to form stars. The gas will also be permanently removed from the galaxy if it is incorporated in other collapsed structures which are not part of the galaxy.

This raises the issue of how critical the original assumption, that the energy released by supernovae is exactly that required to eject the gas to infinity, is to the model. This can be tested by relaxing the original assumption and allowing the ejected gas to have a non-zero energy at infinity. This amount of energy can be treated as a new parameter in the model, positive in cases when the supernovae release more energy than is necessary to unbind the gas and negative in cases where the gas is expelled, but the injected supernovae energy is less than the binding energy.

It is still assumed that once supernovae have produced sufficient energy, all the remaining gas is ejected. This is a simplification, since in reality it is likely that some gas is removed and some stays behind. In all cases the gas takes part in the next hierarchical collapse of a block which contains it and the energy of the gas is preserved as binding energy until that collapse. Gas which is expelled with an energy less than the binding energy may fall back or may be absorbed by surrounding structures, but in both cases it is assumed to remain as gas. It does not form stars or baryonic dark matter.

2.4 Details of the Modification

2.4.1 The original model

Star formation in the original model continues until either all the catastrophically cooled gas has turned into stars or until the energy released by supernovae is sufficient to expel the rest of the gas to infinity. Assuming that in order to form one type II supernova a total mass, $M_{\rm SN}$, of gas has to be turned into stars and that only the gas within radius $R_{\rm CF}$ is available for star formation, the total number of supernovae formed equals $f_{\rm u}\frac{M_{\rm gas}(R<R_{\rm CF})}{M_{\rm SN}}$, where $f_{\rm u}$ is the fraction of the cooled gas that turns into stars. As explained previously, $R_{\rm CF}$ is the radius where the cooling time of the gas equals the free fall time and gas outside this radius remains too hot to turn into stars.

Not all the energy released by supernovae can be used to drive the remaining gas out of the galaxy. As mentioned previously, the interaction of supernovae with the interstellar medium is complex and in the dense regions behind the shock front which results from a supernova explosion, energy can be radiated

away very efficiently. Estimates of both the total energy released by supernovae and the proportion of this energy which is released as kinetic energy differ. The value which is usually assumed is 10^{51} ergs, [29], but there are some indications that this may be overestimated [130]. NF take the amount of supernova energy available to expel the gas from the galaxy as $4 \times 10^{50} \epsilon_{\rm SN}$ ergs, where $\epsilon_{\rm SN}$ can be regarded as an efficiency factor which depends on the density and the metallicity of the gas, amongst other things. This brings the total amount of supernova energy available for expulsion of the gas to $f_{\rm u} \frac{M_{\rm gas}(R<R_{\rm CF})}{M_{\rm SN}} 4 \times 10^{50} \epsilon_{\rm SN}$.

In the NF model this amount of energy is equated to the binding energy of the remaining gas, i.e. the energy required to take the gas to infinity. It is assumed that self gravitation of the gas can be ignored. The dark matter follows a density profile $\rho \propto r^{-2}$ and the gas follows a density profile $\rho \propto r^{-2\beta}$. The catastrophically cooled gas has negligible thermal energy and is assumed to be homogeneous. It has a binding energy of $(1 - f_{\rm u}) 4\sigma^2 M_{\rm gas}(R < R_{\rm CF})$, where $(1 - f_{\rm u}) M_{\rm gas}(R < R_{\rm CF})$ is the amount of cooled gas which has not turned into stars.

The hot gas has a thermal energy of $\frac{3}{2}\sigma^2 \beta M_{\rm gas}(R > R_{\rm CF})$. Its binding energy is written as $\frac{5}{2}\sigma^2 \chi(\beta) M_{\rm gas}(R > R_{\rm CF})$ where $M_{\rm gas}(R > R_{\rm CF})$ is the mass of the hot gas and $\chi(\beta) = (2\beta - 1)(9 - 4\beta)/[5\beta(3 - 2\beta)]$ is a correction factor to account for the difference between the gas temperature and the virial temperature. $\chi(\beta)$ equals one when β is one. If $\beta \leq 0.5$ the total energy of the gas is positive, which means that the gas is not gravitationally bound. If this is the case the gas does not take part in the collapse, and is preserved until a later stage of the collapse hierarchy.

The fraction of the catastrophically cooled gas which turns into stars in the original model, f_u, is calculated by solving

$$f_u \frac{M_{gas}(R < R_{CF})}{M_{SN}} 4 \times 10^{50} \epsilon_{SN} = \frac{5}{2}\sigma^2 \chi(\beta) M_{gas}(R > R_{CF}) + (1 - f_u) 4\sigma^2 M_{gas}(R < R_{CF}) \quad (2.1)$$

The left hand side of this equation gives the energy released by the supernovae, which is available to expel the gas from the galaxy and the right hand side gives the binding energy of the remaining gas, i.e. the gas which has not yet turned into stars. If f_u is calculated to be larger than 1, then the energy released by supernovae is insufficient to unbind the rest of the gas and all the catastrophically cooled gas turns into stars.

2.4.2 The modification

For this work, the fraction of the catastrophically cooled gas that turns into stars, f_*, is no longer precisely that required to produce enough supernovae to unbind the remaining gas. Instead, the energy released by supernovae is taken to be equal to the energy needed to unbind the rest of the gas and to give it an extra amount of specific energy, $\mu\sigma^2$. μ is a new parameter which can be either positive or negative.

The fraction of the catastrophically cooled gas which turns into stars is now calculated by solving

$$f_* \frac{M_{gas}(R < R_{CF})}{M_{SN}} 4 \times 10^{50} \epsilon_{SN} = (\frac{5}{2}\sigma^2 \chi(\beta) + \mu\sigma^2) M_{gas}(R > R_{CF}) + (1 - f_*)(4\sigma^2 + \mu\sigma^2) M_{gas}(R < R_{CF}) \quad (2.2)$$

for f_*. The left hand side of the equation again gives the energy released by supernovae, which is availabe to eject the rest of the gas. The first term on the right hand side gives the total energy released into the hot gas and the second term the total energy released into the remaining cold gas. Solving this yields

$$f_* = \frac{1 + [\frac{5}{2}\chi(\beta) + \mu]M_{\rm gas}(R > R_{\rm CF})/[(4+\mu)M_{\rm gas}(R < R_{\rm CF})]}{1 + 4 \times 10^{50}\epsilon_{SN}/[M_{\rm SN}(4+\mu)\sigma^2]} \tag{2.3}$$

Depending on the values of f_* and $f_{\rm u}$, there are four possibilities. Both f_* and $f_{\rm u}$ can be smaller than 1, $f_{\rm u} < 1$ but $f_* > 1$, $f_{\rm u} > 1$ but $f_* < 1$ and both f_* and $f_{\rm u}$ larger than 1. Each of these possibilities relates to a different physical situation.

If both f_* and $f_{\rm u}$ are smaller than one, a fraction, f_*, of the catastrophically cooled gas turns into stars and the remaining gas is expelled. For positive values of μ the fraction which turns into stars is larger in the modified model and the remaining gas ends up with more energy. For negative values of μ, less gas turns into stars and the unbound gas will contain less energy. In both cases a dwarf galaxy is formed.

If $f_{\rm u} < 1$ but $f_* > 1$, which can only happen when $\mu > 0$, all the catastrophically cooled gas is turned into stars and the rest of the gas becomes unbound. This situation is not fundamentally different from the first case when $\mu > 0$, except that the total amount of star formation is limited by the amount of catastrophically cooled gas. Again, in both the modified and the unmodified model a dwarf galaxy is formed.

A more complicated case occurs when $f_{\rm u} > 1$ but $f_* < 1$, which can happen if $\mu < 0$. In the unmodified model this means that all the catastrophically

cooled gas turns into stars and no gas is expelled from the galaxy. However, in the revised model some catastrophically cooled gas is left after star formation. Since the physical justification for modifying the model is that the gas is either removed from the galaxy until the next collapse, or that it falls back into the galaxy or another structure, but does not form stars, in this case the remaining gas is assumed to be expelled from the galaxy, but with a lower energy than in the original model. This means that, while in the unmodified model any hot gas would form baryonic dark matter and the galaxy might end up as a normal galaxy, in the modified model a dwarf galaxy is formed with no baryonic dark matter. In this case the modification makes a significant difference for the individual galaxy, although this may be masked by results of subsequent collapses. Also, the number of cases where this happens is expected to be small.

Finally, if both $f_{\rm u}$ and f_* are larger than 0, all the catastrophically cooled gas turns into stars and no gas is expelled from the galaxy. This is occurs in collapses of the larger protogalaxies, which produce significant amounts of hot gas. It is the case where the modification makes no difference.

Chapter 3

Results of the Modified Feedback Rule

3.1 The simulations

Both the model described by NF97 and the model described by NF00 were modified as described in section 2.4.2. This adds the new model parameter μ, such that the gas expelled from a galaxy by supernova explosions has net energy, i.e. thermal energy, kinetic energy and gravitational potential energy relative to the galaxy, of $M_{\rm gas}\mu\sigma^2$. Simulations were run with values for μ varying from -0.2 to 1.

3.2 Results for the NF97 Model

In most respects the effect of the modification is small. The main exception is that the fraction of early type (elliptical) galaxies is reduced. Although the modification has a direct influence on the amount of star formation in individual collapses, the total amount of star formation remains virtually unchanged (Table 3.1).

μ	total star formation ($10^{12}M_\odot Mpc^{-3}$)
-0.2	2.43×10^{-3}
-0.1	2.44×10^{-3}
0	2.45×10^{-3}
0.1	2.46×10^{-3}
0.2	2.47×10^{-3}
0.4	2.47×10^{-3}
0.6	2.47×10^{-3}
0.8	2.46×10^{-3}
1.0	2.45×10^{-3}

Table 3.1: Total star formation

This is remarkable as the star formation in dwarf galaxies can change up to 25%. The parameters of the Faber-Jackson relation are also independent of μ within

μ	M_* ($10^{12}M_\odot$)	α	n_* ($10^{-3}Mpc^{-3}$)
-0.2	0.145	-1.26	4.63
-0.1	0.143	-1.26	4.53
0	0.142	-1.27	4.35
0.1	0.143	-1.28	4.16
0.2	0.143	-1.28	3.99
0.4	0.134	-1.29	3.87
0.6	0.129	-1.30	3.64
0.8	0.128	-1.31	3.33
1.0	0.129	-1.32	3.07

Table 3.2: Schechter function parameters.

error margins.

The Schechter function is a good fit to the galaxy luminosity function for all values of μ. Plots of the Schechter function for $\mu = 0$ (Figure 3.1) and for $\mu = 1$ (Figure 3.2) are shown. The Schechter function parameters for different values of μ are given in Table 3.2. Increasing μ causes a decrease in M_* of up to 10% and minor changes in α. The main change is in n_*, which decreases up to 30%. Some possible explanations are discussed later in this section.

Looking at cluster properties, it can be seen that the mean gas fraction and

μ	cluster mean gas fraction	cluster mean abundance
-0.2	0.143	0.215
-0.1	0.142	0.215
0	0.142	0.215
0.1	0.141	0.215
0.2	0.141	0.214
0.4	0.140	0.211
0.6	0.140	0.207
0.8	0.139	0.205
1.0	0.139	0.203

Table 3.3: Cluster mean gas fraction and cluster mean abundance.

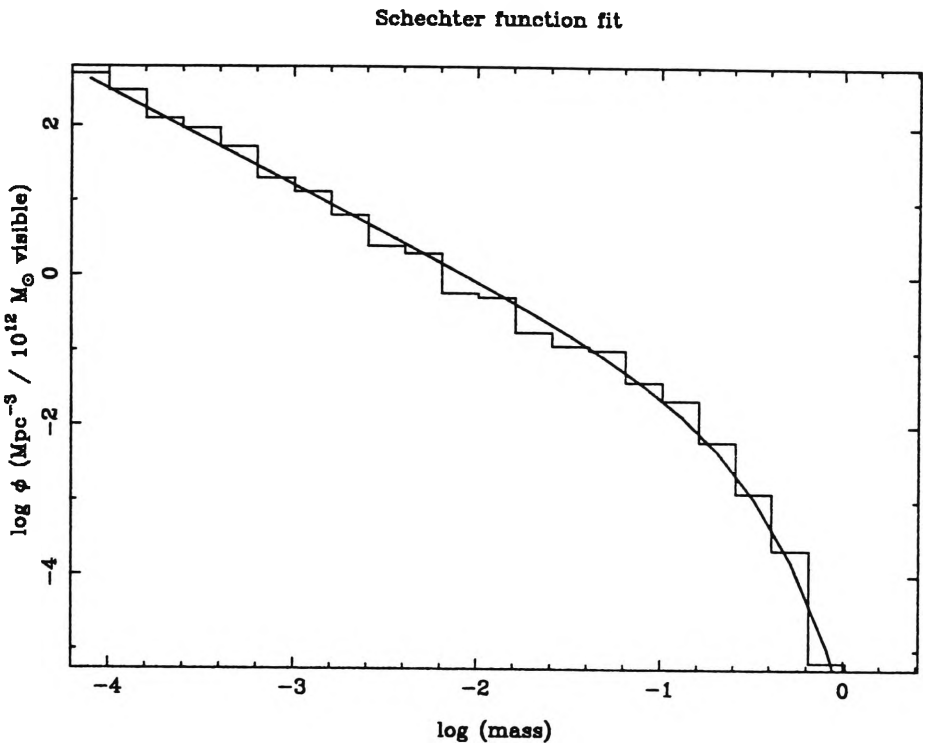

Figure 3.1: Schechter function for $\mu = 0$

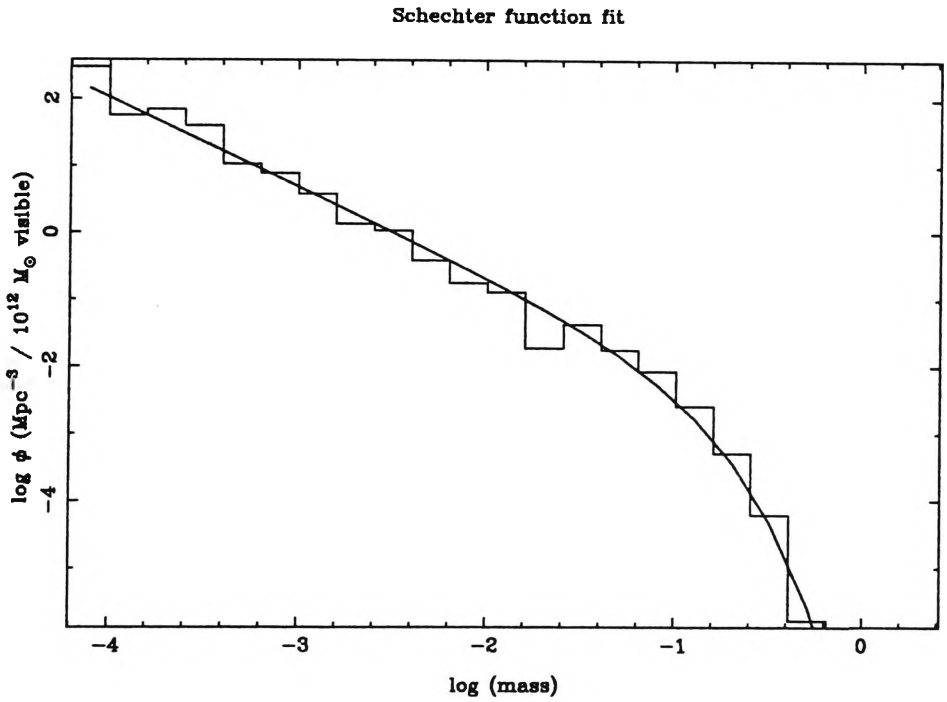

Figure 3.2: Schechter function for $\mu = 1.0$

abundance (Table 3.3) decrease but these changes are only a few percent. The cluster $\beta's$ also get slightly lower, which means the gas in clusters is less tightly bound, but the change is minimal.

While the total star formation remains constant, the history of star formation does change with μ. Figures 3.3 and 3.4 show star formation as a function of redshift. Although the change is modest, it can be seen that for higher values of μ star formation shifts to higher redshifts. It should be noted that the model only includes spheroid star formation. Star formation in discs is not included. Plots of star formation as a function of μ (Figure 3.5 and Figure 3.6) show that star formation increases with μ for high redshifts, but decreases with μ for low redshifts, with the turning point around $z = 4$. These two effects tend to cancel each other with the result that the total star formation remains unchanged.

The shift of star formation from lower to higher redshifts can be understood as follows. In small dwarf galaxies, where star formation continues until the supernovae have released sufficient energy to expel the remaining gas from the galaxies, star formation continues longer for positive values of μ, since more supernovae are required to supply the energy to eject the gas. The number of stars that are formed in small dwarf galaxies therefore increases with increasing μ. This has two consequences for subsequent collapses: the total gas fraction decreases (since more gas has already formed into stars) and the energy content of the remaining gas increases. The result is that, for later collapses, $R_{\rm CF}$, the radius within which the gas can cool catastrophically, moves inward. This means a reduction in the amount of gas available for star formation. If there is less star formation, there will also be fewer supernovae, and at an earlier stage the

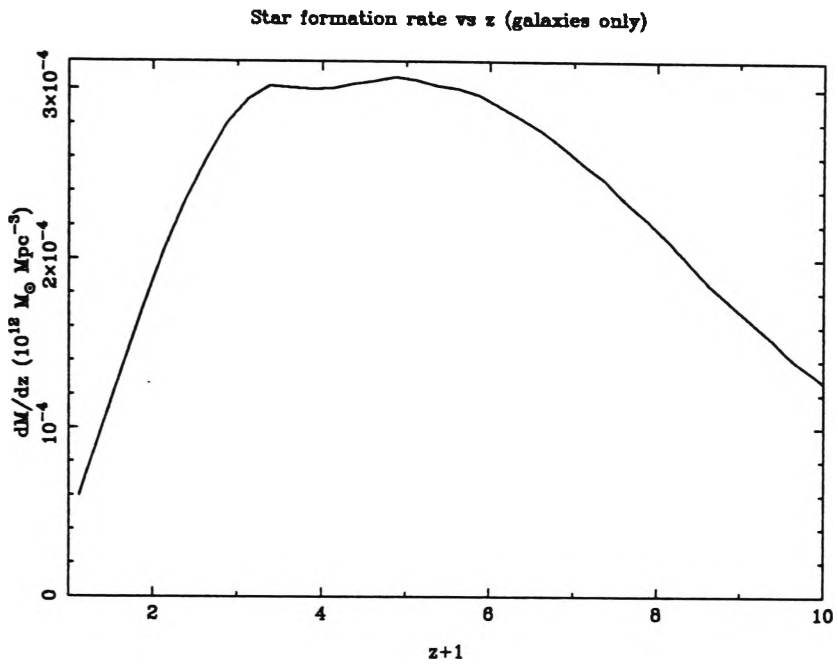

Figure 3.3: Star formation for μ=0

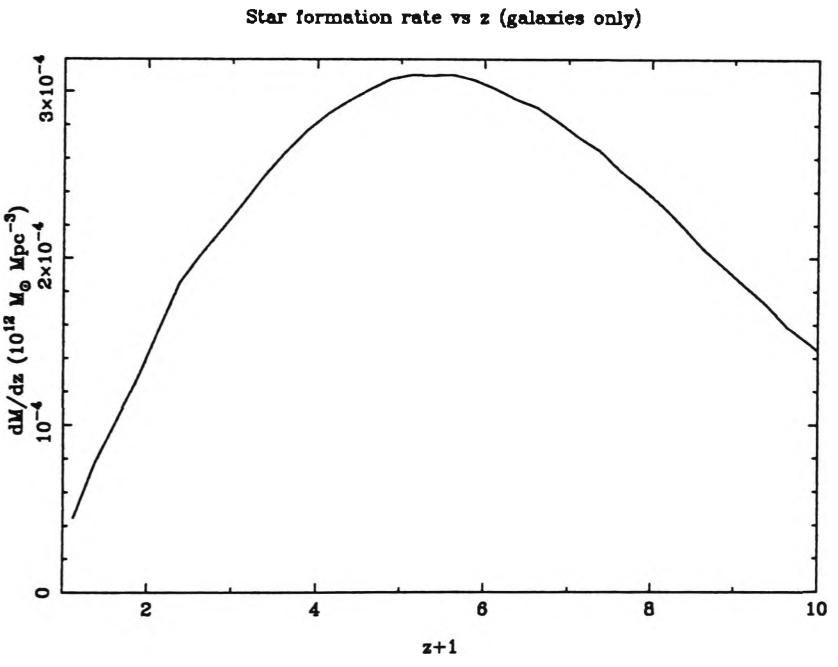

Figure 3.4: Star formation for μ=1.0

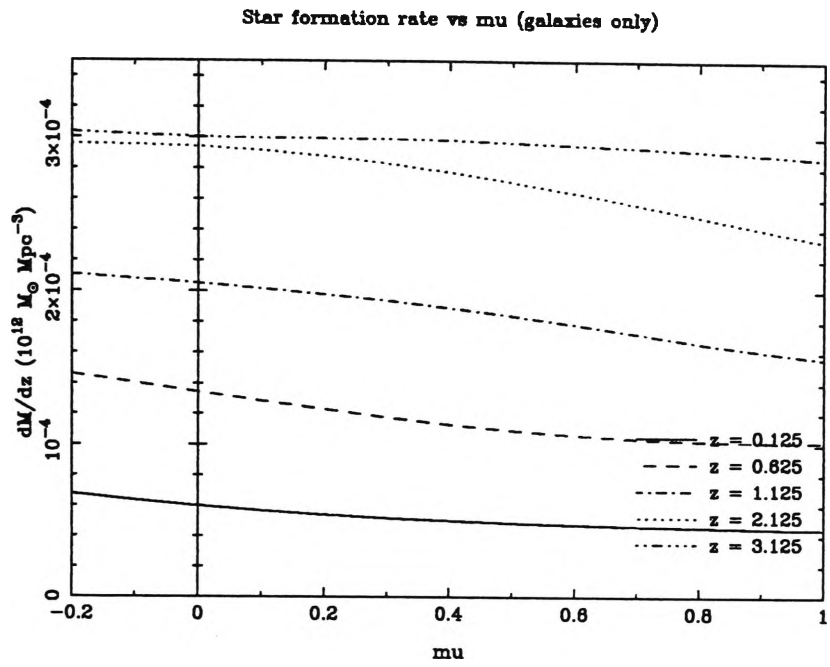

Figure 3.5: Star formation vs μ

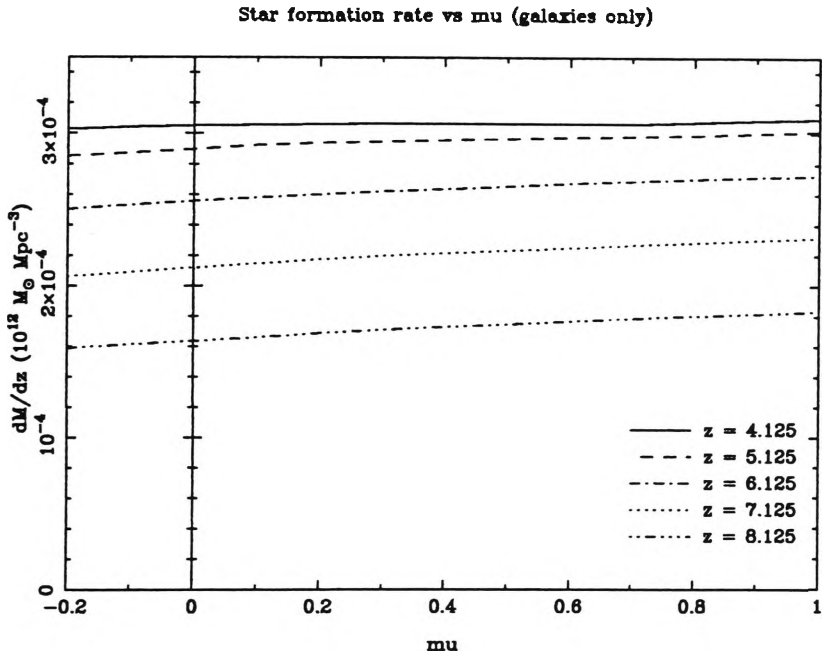

Figure 3.6: Star formation vs μ

energy released by supernovae is no longer be sufficient to eject the remaining gas from the galaxy. As a result, the transition from dwarf galaxies to normal galaxies takes place earlier, at higher redshifts. As explained in chapter 2, once supernovae can no longer drive the remaining gas out of the galaxy, increased values of μ no longer result in increased star formation. All the catastrophically cooled gas in these objects turns into stars, independent of the value of μ. The combination of the lower gas fraction and the higher energy content of the gas, however, results in a smaller amount of gas that can cool catastrophically. The consequence of this is a decreased star formation rate at later times.

It is now possible to explain the change in the Schechter function as well. The Schechter function is fitted to the combined numbers of dwarf and normal elliptical galaxies. An earlier transition from dwarf to normal galaxies reduces the total number of dwarf galaxies at the high mass end of the dwarf distribution. As will be discussed below, many of these former dwarfs are turned into normal disc galaxies. Less star formation at later times also reduces the number of very luminous elliptical galaxies. This depletes the high mass end of the Schechter function, resulting in a decrease in M_*. The combined reduction in the numbers of large dwarf galaxies and normal elliptical galaxies causes the reduction in n_*.

As mentioned above, the most significant influence of non-zero μ is on the proportions of the different morphological types of galaxy. The fraction of elliptical galaxies, both in the field and in clusters, decreases more than 30% when μ increases from 0 to 1.0 (Figure. 3.7). This is not surprising, since galaxy morphology was found to be sensitive to model parameters by NF97.

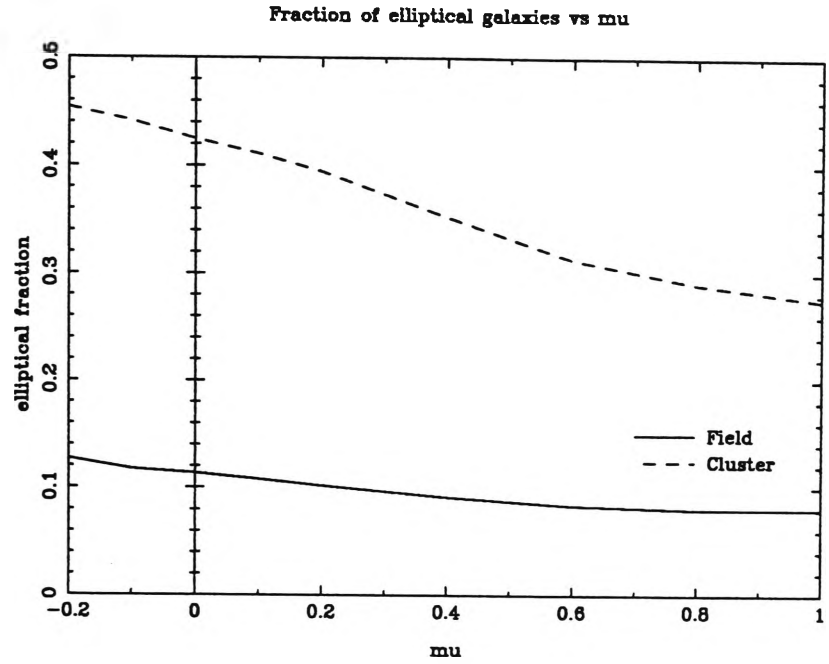

Figure 3.7: Fraction of elliptical galaxies

In the NF model discs form when the last of the hot gas cools. The disc formation rate (a crude estimate of the star formation rate in discs), as shown in Figure 3.8 for $\mu = 0$ and in Figure 3.9 for $\mu = 1.0$, is found by taking the whole of each hot halo which finishes cooling, to be turned into a disc. Since only a small part of this gas (perhaps 10%) actually forms the disc, this overestimates the disc formation rate. The real disc formation rate is about an order of magnitude lower.

The disc formation rate shows a complicated dependence on μ. For higher values of μ a double peak occurs, the earlier one at about $z = 3$ the later one at $z = 0.7$. Plots of the disc formation rate vs μ (Figure 3.10 and Figure 3.11) show that, for lower redshifts the disc formation rate increases with μ for lower values of μ but decreases with μ for higher values of μ. For higher redshifts this behaviour is inverted, but at the highest redshifts the formation rate increases

monotonically with μ.

There is no simple explanation for this as there are several competing effects which influence the disc formation rate. In the model, a disc galaxy is formed if all the hot gas is able to cool before the next collapse. Whether this is possible depends on the temperature and density of the gas, and on the time available until the next collapse. Higher values of μ cause an earlier transition to normal galaxies, as described above, and therefore result in more normal galaxies with lower masses. The temperature of the gas in these galaxies is low, making its cooling time shorter and so it is likely that the gas will be able to cool before the next collapse. This is the first effect. The second effect is the decreased gas fraction as the result of more gas being turned into stars in earlier collapses. This reduces the density, which means that the cooling time increases and less gas can cool before the next collapse. The third effect is caused by the increased energy content of the gas. This raises the temperature, increasing the cooling times which means there is a greater chance that there will be another collapse before all the gas has cooled. Fourth, at earlier times, there is generally less time between collapses, so that the earlier normal galaxies typically have less time to cool and form a disc galaxy. The first effect increases the disc formation rate, resulting in a larger number of spiral galaxies, the second, third and fourth effects counteract this, tending to increase the fraction of elliptical galaxies.

The competition between these effects is complicated by further factors, such as the temperature and metallicity dependence of the cooling function. As Figures 3.10 and 3.11 illustrate, which ones dominate depends subtly on the values of μ and the redshift. The increase in total disk formation not only results in a

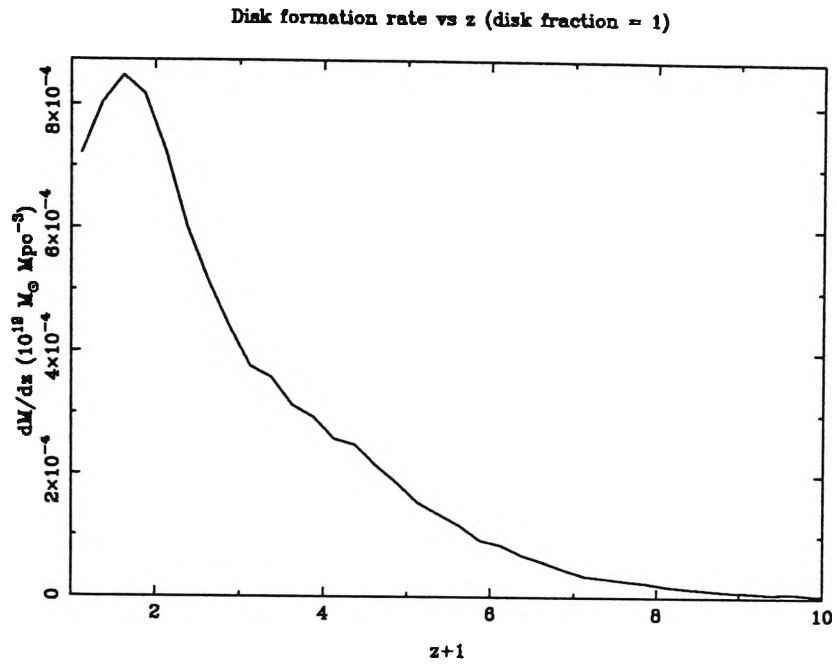

Figure 3.8: Disc formation for $\mu = 0$

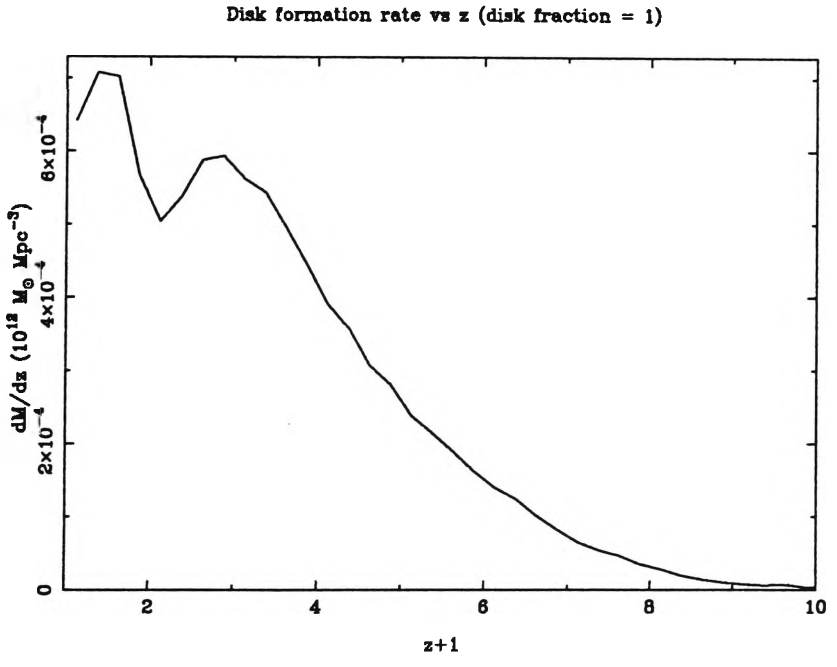

Figure 3.9: Disc formation for $\mu = 1.0$

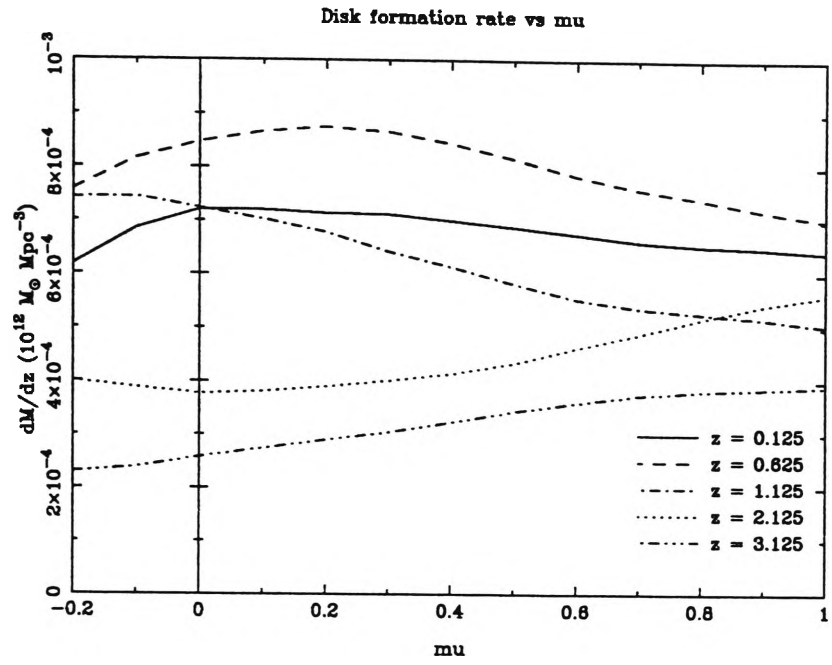

Figure 3.10: Disc formation vs μ

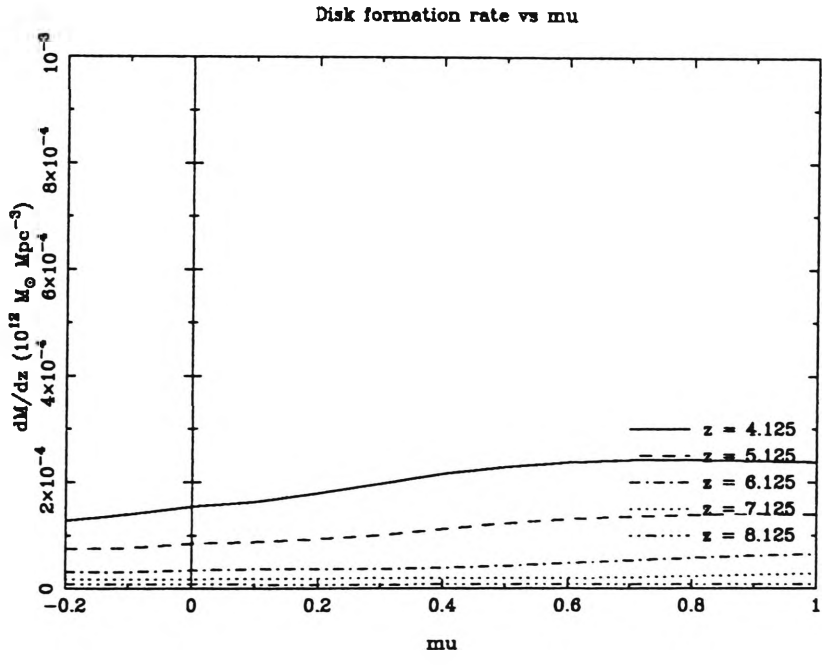

Figure 3.11: Disc formation vs μ

reduced fraction of elliptical galaxies but also has an influence on cluster properties. According to the model, gas which does not turn into stars but is able to cool before the next collapse is transformed into baryonic dark matter. Because there is more star formation during the earlier collapses, the gas which is transformed into baryonic dark matter has a relatively high metal abundance. These metals are permanently lost. Less star formation during subsequent collapses then results in a reduced abundance for the cluster gas. In other words, the total amount of metals released by supernovae remains unchanged but a larger fraction of these metals ends up in baryonic dark matter.

3.3 Results for the NF00 Model

The main difference between this and the earlier model is the low density cosmology ($\Omega = 0.3$). In this model the influence of μ is even smaller than in the model with $\Omega = 1$. Again the total amount of star formation remains remarkably constant (Table 3.4).

μ	total star formation ($10^{12} M_\odot Mpc^{-3}$)
-0.2	3.25×10^{-4}
-0.1	3.28×10^{-4}
0	3.33×10^{-4}
0.1	3.32×10^{-4}
0.2	3.34×10^{-4}
0.4	3.37×10^{-4}
0.6	3.38×10^{-4}
0.8	3.40×10^{-4}
1.0	3.40×10^{-4}
1.5	3.33×10^{-4}
2.0	3.24×10^{-4}

Table 3.4: Total star formation.

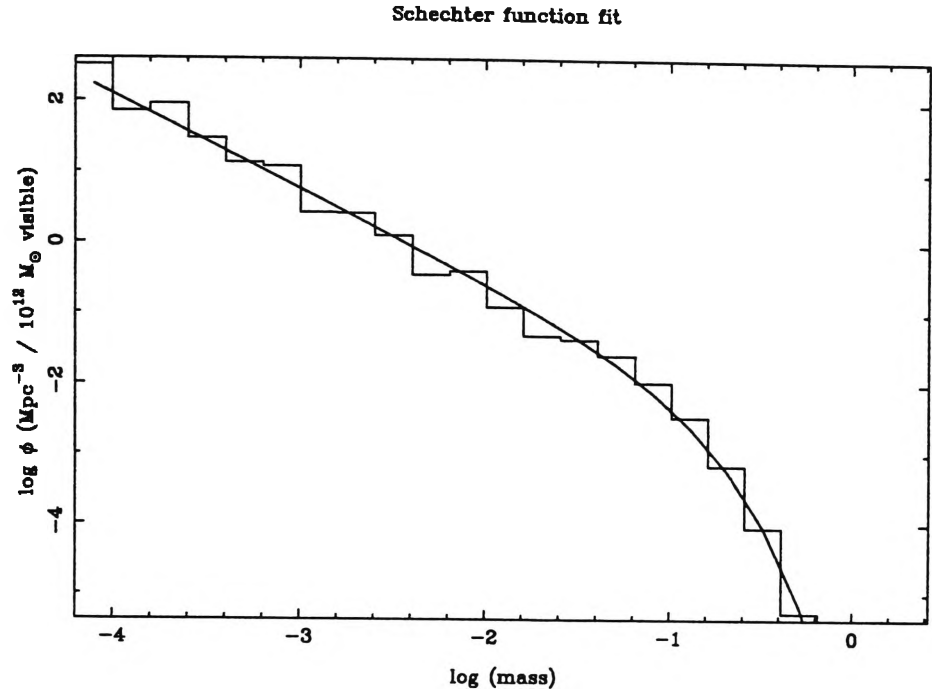

Figure 3.12: Schechter function for $\mu = 0$

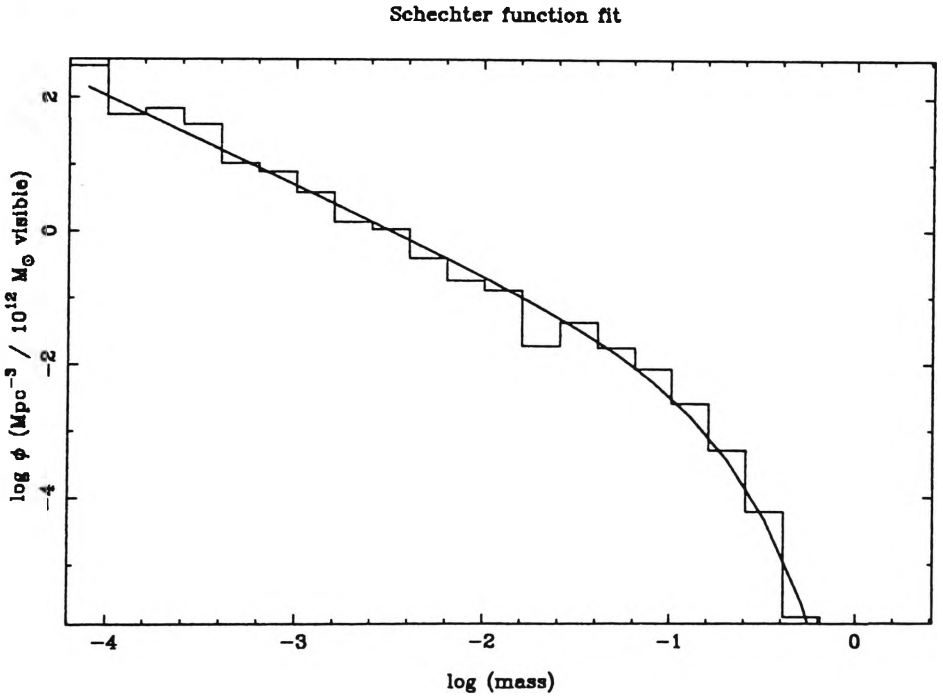

Figure 3.13: Schechter function for $\mu = 1.0$

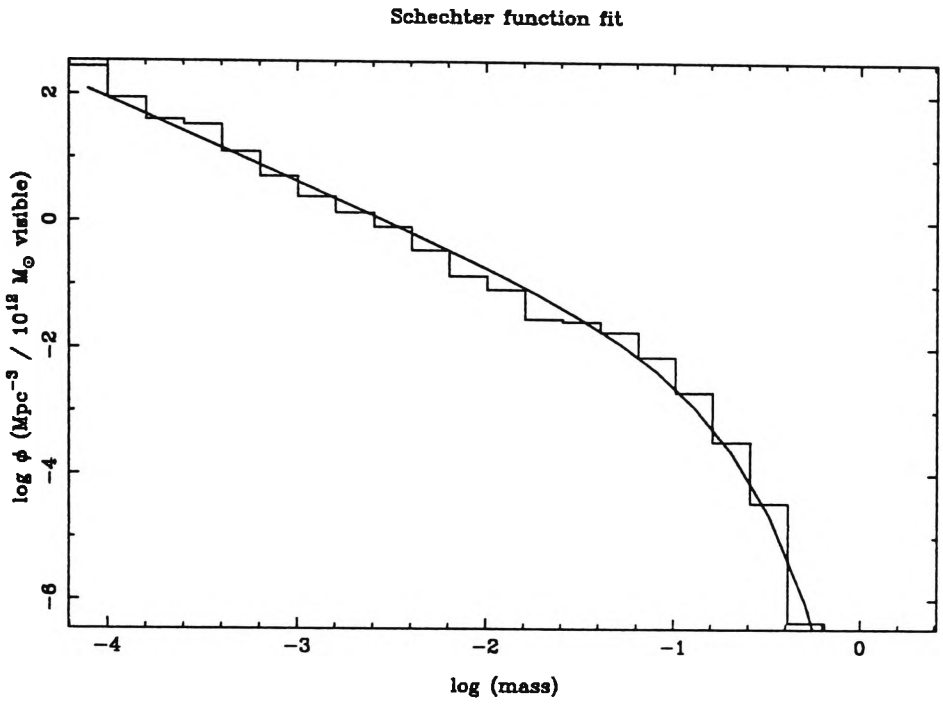

Figure 3.14: Schechter function for $\mu = 2.0$

μ	M_* ($10^{12}M_\odot$)	α	n_* (10^{-3}Mpc^{-3})
-.2	.088	-1.29	1.62
-.1	.095	-1.32	1.37
0.	.096	-1.32	1.30
.1	.092	-1.32	1.37
.2	.090	-1.32	1.37
.4	.087	-1.31	1.36
.6	.087	-1.32	1.25
.8	.088	-1.33	1.14
1.	.084	-1.33	1.13
1.5	.072	-1.30	1.22
2.	.071	-1.30	1.17

Table 3.5: Schechter function parameters.

The Schechter function is also a good fit to the galaxy luminosity function for all values of μ (Figure 3.12 and Figure 3.13) . Even when the value of μ is increased to 2., which is almost certainly unphysical, the Schechter function still fits the data easily(Figure 3.14). The Schechter function parameters (Table 3.5) show a slight decrease in M_* and a 30 % variation in n_*, as in the other model.

As in the first model, increasing μ results in a slight decrease in cluster gas fractions (Table 3.6), although this is barely significant. The abundance, however, is constant within error margins. Cluster $\beta's$ ($\beta = m\sigma^2/kT_{\rm gas}$) show no major differences.

μ	cluster mean gas fraction	cluster mean abundance
-0.2	0.127	0.144
-0.1	0.126	0.145
0	0.123	0.144
0.1	0.125	0.145
0.2	0.125	0.146
0.4	0.124	0.147
0.6	0.123	0.148
0.8	0.123	0.148
1.0	0.122	0.148
1.5	0.122	0.146
2.0	0.123	0.143

Table 3.6: Cluster mean gas fraction and cluster mean abundance.

Although the history of star formation varies in a similar manner to the way it does in the NF97 model, the effect is smaller. Plots of star formation for $\mu = 0$ (Figure 3.15) and for $\mu = 1$ (Figure 3.16) show modest differences. As before, star formation increases with increasing values for μ at high redshifts, but decreases with increasing values for μ at low redshifts (Figure 3.17 and Figure 3.18). The turning point is at a lower redshift than in the first model, at $z \approx 2$ instead of

$z = 4$. Changes are at most a few percent, so it is not surprising that the total star formation remains unchanged.

As explained in the previous section, the influence of μ on the star formation rate is the result of two competing effects. The first effect is that increasing μ results in more star formation in small dwarf galaxies, because star formation continues after the amount of energy released by supernovae is sufficient to unbind the remaining gas. The second effect is that the reduced gas fraction and increased energy content of the gas during subsequent collapses result in less star formation at later times. These two effects can be seen in both models. In the second model, however, there is relatively less star formation at higher redshifts than in the first model (compare Figures 3.3 and 3.15). This means that although the modification has much the same effect on early collapses in both models, its absolute effect is much smaller in the second model. Since the second effect, the reduced star formation rate at later times, is only a consequence of the first effect during earlier collapses, the reduction in the star formation rate is small at intermediate redshifts, and the first effect dominates for a longer time. This explains the fact that the transition from an increase to a decrease in star formation with increasing values of μ takes place at a lower redshift in the second model.

As in the first model, the most significant effect of the modification is on the ratio of elliptical to spiral galaxies (Figure. 3.19). The fraction of elliptical galaxies in clusters decreases by 20% when μ is increased from 0 to 1. For the same values of μ the fraction of elliptical galaxies in the field decreases by 40%. The disc formation rate (Figure 3.20, Figure 3.21, Figure 3.22) shows a slight decrease with increasing μ for redshifts below 0.5, an increase for redshifts between 0.5

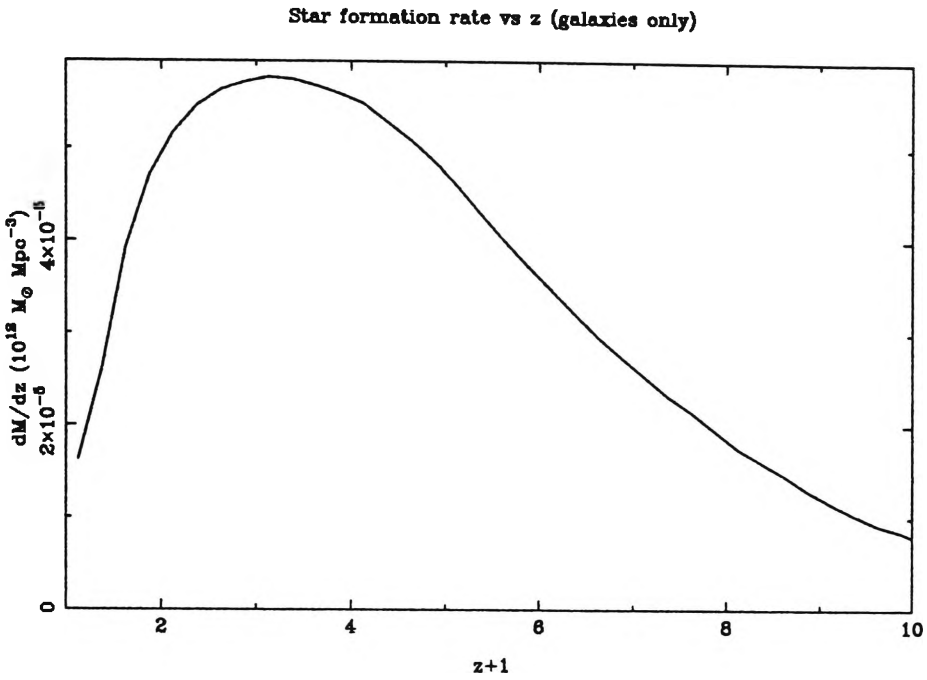

Figure 3.15: Star formation for μ=0

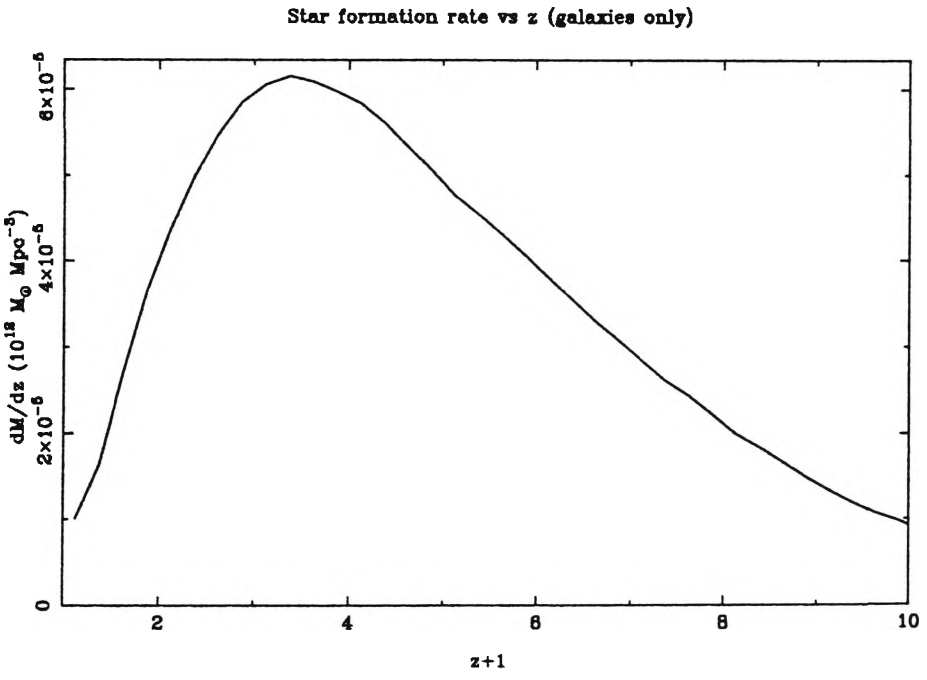

Figure 3.16: Star formation for μ=1.0

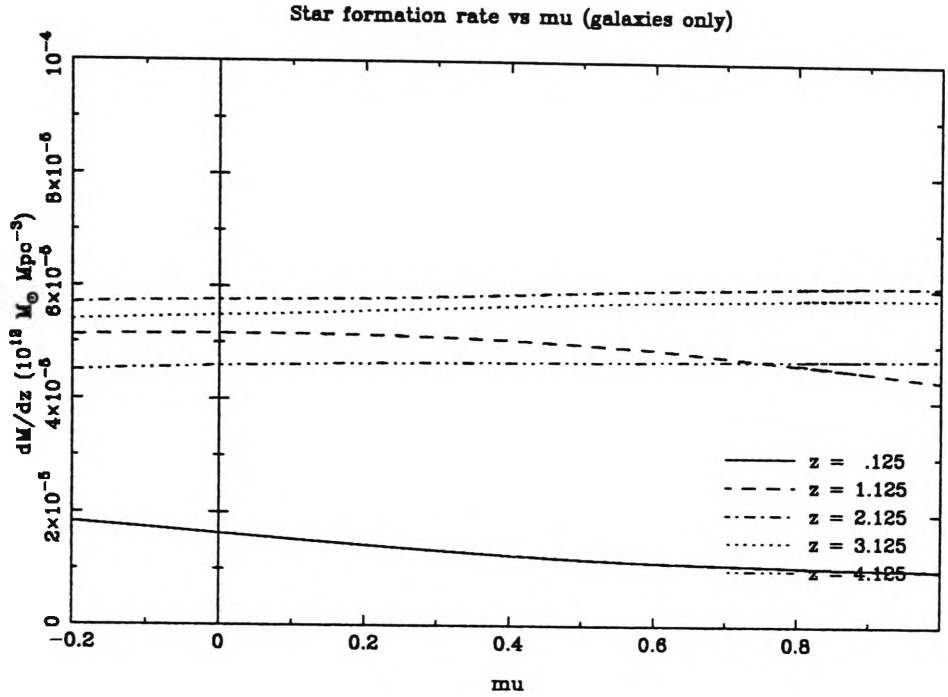

Figure 3.17: Star formation vs μ

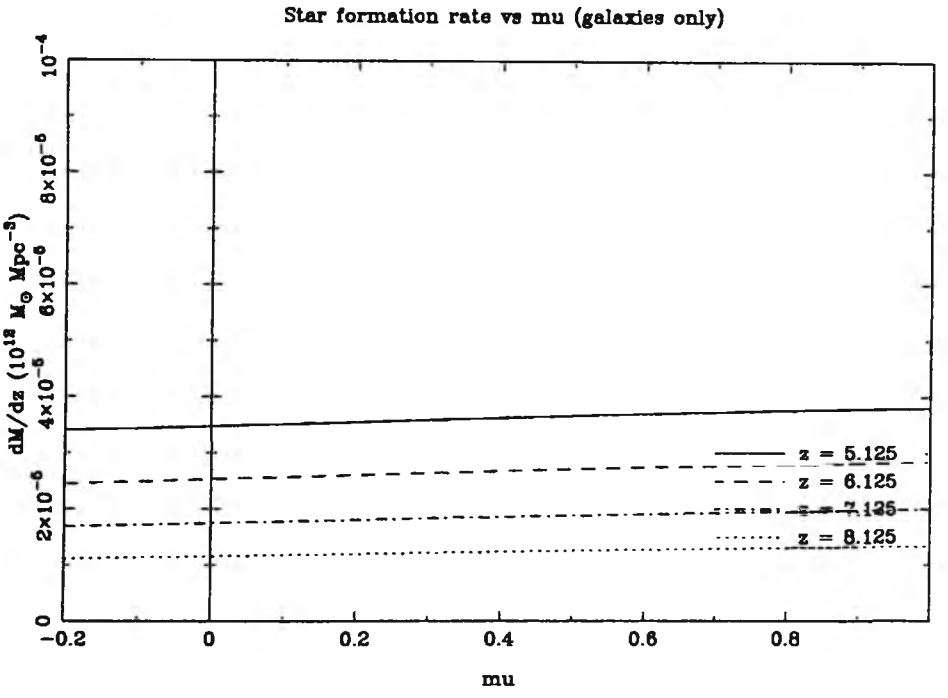

Figure 3.18: Star formation vs μ

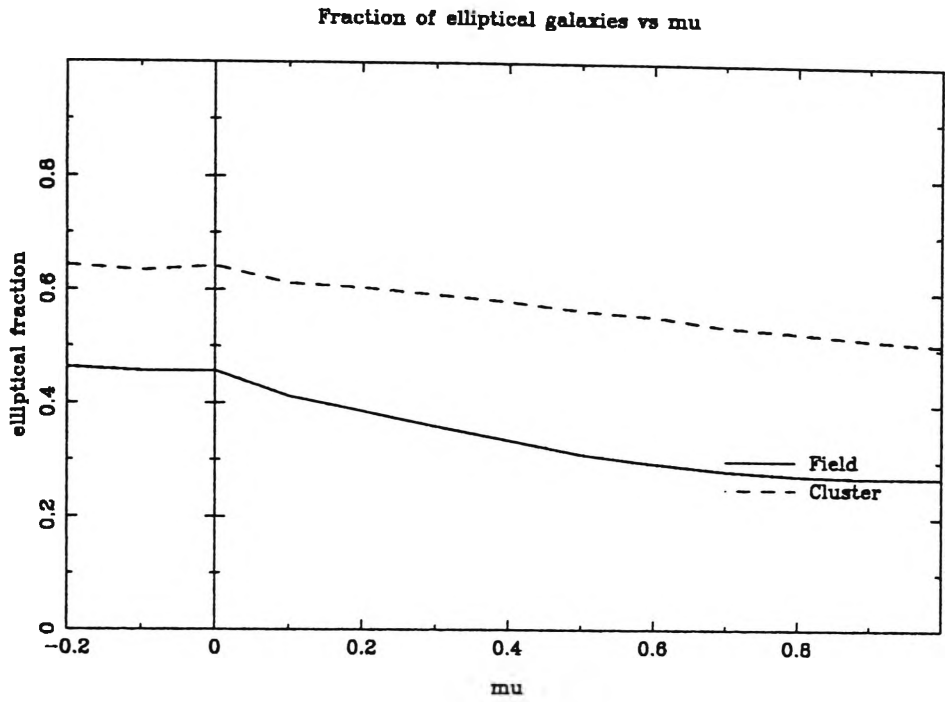

Figure 3.19: Fraction of elliptical galaxies

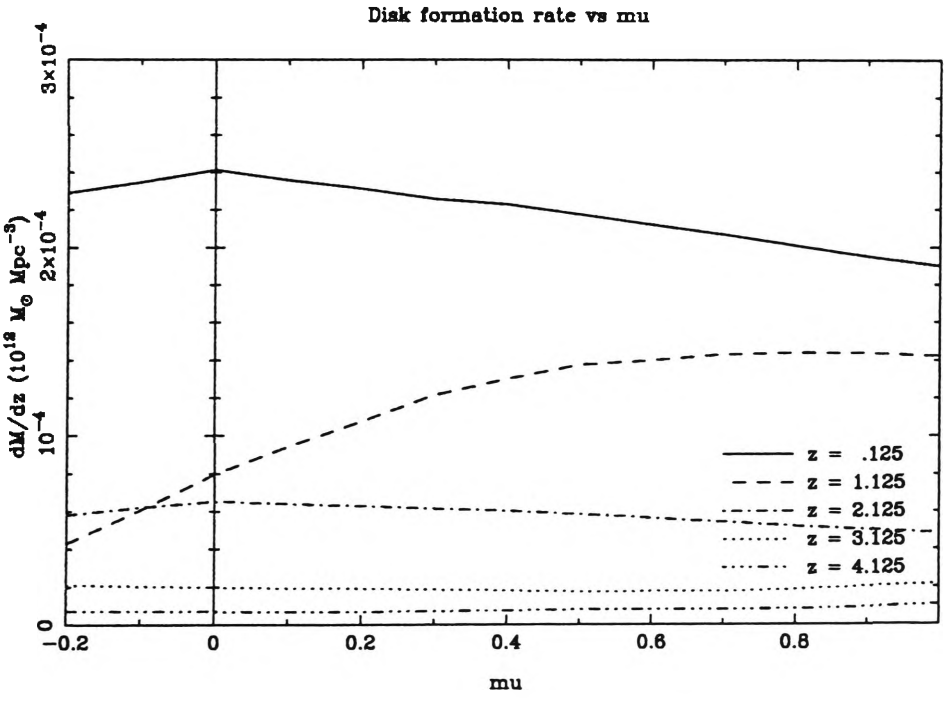

Figure 3.20: Disc formation vs μ

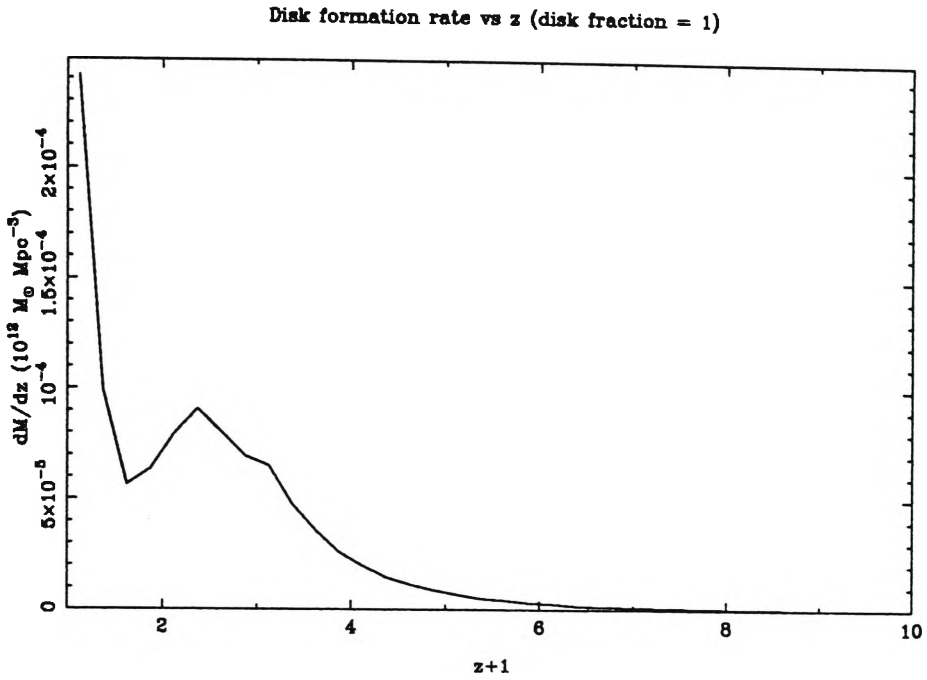

Figure 3.21: Disc formation for $\mu = 0$

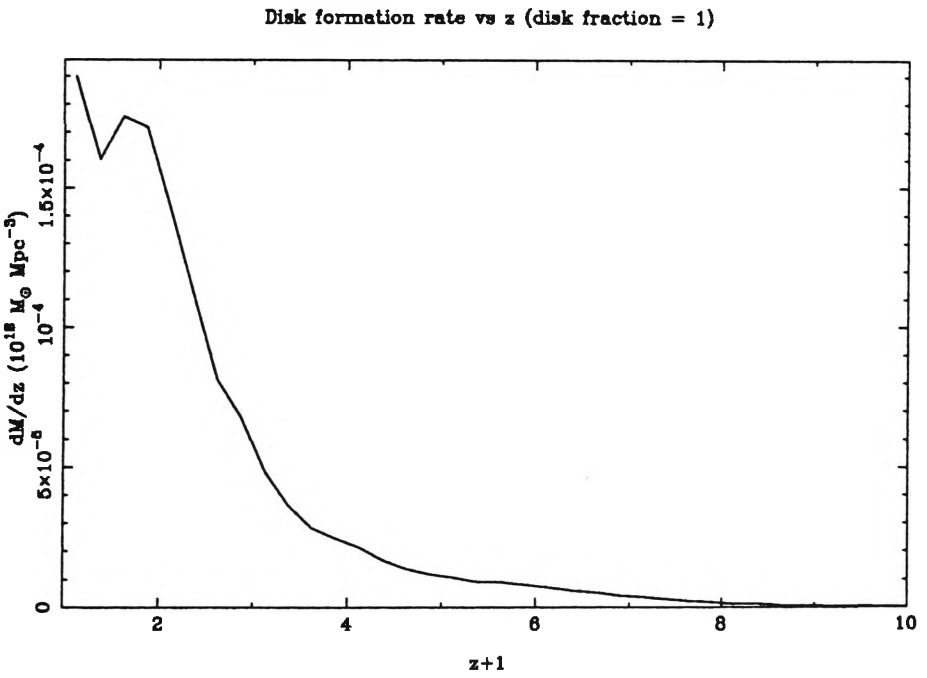

Figure 3.22: Disc formation for $\mu = 1.0$

and 2 and little change above that. The dominant effect here is probably the earlier transition from dwarf to normal galaxies.

3.4 Conclusions

In general the results of these simulations are not sensitive to the assumption being tested, that total star formation proceeds until either all the catastrophically cooled gas is turned into stars or the amount of energy released by supernovae precisely equals the binding energy of the remaining gas. The outcome that is most sensitive to this assumption is the ratio of elliptical to spiral galaxies, but this is known to be sensitive to other model parameters. Most other properties remain within error margins as the parameter μ varies from -0.2 to 1. Total star formation remains remarkably constant as μ is varied. This is the result of two competing effects. There is more star formation during early collapses which depletes the remaining gas and increases its energy content, reducing the star formation resulting from subsequent collapses. The greater the first effect during early collapses, the greater is the second effect during later collapses. As a result, the total star formation is very insensitive to the value of μ. This means that the amount of star formation that occurs in dwarf galaxies before the remaining gas is expelled from the galaxy does not influence the total star formation. It should be noted that these statements only apply to star formation in spheroids. Since the fraction of spiral galaxies increases with increasing μ, star formation is discs could be sensitive to μ.

The reduction of the fraction of elliptical galaxies with increasing μ is its most significant effect. Three factors which influence the disc formation rate and which

are affected by changing μ are: the mass at which the transition from dwarf to normal galaxy occurs, the gas fraction and the energy content of the gas.

The effect of modifying the collapse model is qualitatively similar in high and low density cosmological models although there are quantitative differences. In the low density model, most effects are smaller and the transition from increased to decreased star formation rates with increasing values of μ takes place at a lower redshift. The reduction in the elliptical fraction with increasing values of μ is similar in the two models but this fraction is higher in the low density model for all values of μ.

The objective of this work was to determine if the NF model is sensitive to one of its weakest assumptions. The results here show that it is not sensitive to modest changes in that assumption. The model outcome that is most sensitive to the assumption, galaxy morphology, is more sensitive to several other model parameters, so that it is not made significantly less certain by this assumption. It can be concluded that the assumption made by Nulsen and Fabian, that total star formation proceeds until either all the catastrophically cooled gas is turned into stars or the amount of energy released by supernovae is just enough to unbind the rest of the gas, has a relatively minor influence on the results of their model.

Chapter 4

N-body Hydrodynamic Codes

4.1 Introduction

N-body methods have been around for several decades. By 1941 a gravitational N-body simulation had already been performed, using an analog optical computer. Gravity was represented by the flux of 37 light bulbs, and photocells and galvanometers were used to measure and display the inverse square law force. Digital computers were first used in the early 1960s and from then on increasingly large simulations were performed. From the initial simulations of less than 100 particles, N-body models have now grown to many millions of particles. [105] [142]

The essence of an N-body model is that matter is represented by a set of particles which move under the influence of gravity. These particles cannot necessarily be identified with physical objects. Their mass is usually set by the mass of the system that is being simulated and the resolution, determined by the available computing resources. While on large scales gravity is the dominant force, gravitational computations alone are not sufficient to simulate the formation of galaxies. As explained in chapter 1, gas processes are an essential part of galaxy formation, and realistic simulations need to include them. N-body codes are therefore often combined with hydrodynamic computations.

Hydrodynamic codes can be divided into two main categories: Eulerian and Lagrangian methods. The difference between these methods is in the derivatives used in the fluid equations. Eulerian methods solve the flow equations using a fixed grid, so that the fluid passes through the Eulerian cells. The flow equations can than be regarded as partial differential equations in position and time with

the convective derivation expressed as

$$\frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla$$

where $\mathbf{u}$ is the flow velocity. In the Lagrangian form of the fluid equations the Lagrangian derivative

$$\frac{d}{dt} = \frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla$$

is used. The Lagrangian derivative of a quantity is the rate of change of a quantity as seen from a point moving with the fluid. In other words Eulerian methods solve the fluid equations relative to a fixed grid in space whereas Lagrangian methods solve the fluid equations relative to the moving fluid.

Eulerian codes have been used in a number of cosmological simulations [112] [28] [27] [111]. Eulerian codes tend to be conceptually simpler, especially in more than one dimension. Because the fluid moves relative to the Eulerian cells they tend to be more prone to numerical diffusion of momentum and heat. Lagrangian codes are generally more stable and deal better with large density differences. They are generally more complex, with the exception of SPH, but it is slower than other codes with about the same resolution. The nature of gravitational instability is such that perturbations on smaller and smaller scales are constantly being generated and the fixed spatial resolution of Eulerian codes prevents the researcher from describing these small scales with sufficient accuracy. To overcome this problem mesh refinement techniques have been applied [4] in which smaller subgrids were constructed to resolve finer scale structures. Bryan and Norman [20] [97] used an adaptive mesh refinement technique. A Lagrangian method using a grid was developed by Gnedin [54] who applied a mesh which deformed

with the flow. This method combines some of the advantages of Eulerian codes with the advantages of Lagrangian methods. In case of severe distortion of the grid, however, the code reverts to an Eulerian scheme.

The most commonly used Lagrangian method in astrophysical simulations is Smooth Particle Hydrodynamics (SPH). SPH is a fully Lagrangian method which follows the fluid variables using particles representing fluid elements. SPH will be discussed in detail in section 4.3. A number of comparisons between SPH and Eulerian codes have been made [66] [8] [83] [117]. Frenk et al. [49] simulated the formation of an X-ray cluster using 12 different codes, including several implementations of SPH and grid based methods with fixed, adaptive and deformable meshes. They concluded that the different methods agreed well, although some properties, like the X-ray luminosity differed by more than a factor of 2.

4.2 Gravitational Force Calculation

The gravitational force on a particle of mass m_j at position $\mathbf{x}_j$ is

$$\sum_i \frac{G m_i m_j}{|\mathbf{x}_i - \mathbf{x}_j|^3}(\mathbf{x}_i - \mathbf{x}_j) \tag{4.1}$$

Calculating the gravitational forces by direct summation of the gravitational forces of all particle pairs is very time consuming as it involved N^2 calculations per step [123]. This severely limits the number of particles that are feasible. Other approximate methods were therefore developed which reduced the number of calculations to $O(NlogN)$ (for an overview see [13]). The most commonly used methods can be divided into two categories: mesh methods and tree codes.

The first mesh method to be developed was the particle-mesh (PM) method. The PM algorithm consists of three steps. First the mass density field is computed on a grid. Then the potential is solved on a grid. Finally the force on the particle is calculated by interpolation. The PM method is fast but the resolution is limited to the resolution of the grid.

Particle-Particle-Particle-Mesh methods P^3M (for a review of PM and P^3M methods see [62]) overcome this limitation by calculating the gravitational forces relative to the grid for distant particles but by direct summation for close particles, i.e. particles less than a few grid spacings away. This greatly improves the accuracy of the force calculations. In case of severe clustering, however, the computational cost of the direct summation becomes too high as many particles are located within a few grid spacings. Couchman [33] [34] developed an adaptive particle-particle-particle-mesh method which places subgrids over regions of high density and only applies pair summation to particles within a few spacings of the subgrid mesh. An alternative approach is multi-resolution mesh method, which are similar to PM method but use mesh refinements where required (e.g. [137] [4]).

The alternative to mesh methods is tree codes [7] [12]. Tree codes replace distant groups of particles by nodes and calculate the force on a particle by summing particle-particle interactions for nearby particles and particle-node interactions for distant particles. Two commonly used methods of constructing the tree are the Barnes and Hut (BH) [7] algorithm and the Benz, Bowers, Cameron and Press (BBCP) [12] algorithm. The difference between these two methods is in the construction of the tree. The BH tree is constructed top down. For a three

dimensional simulation it starts with a cube which is large enough to contain all the particles in the system. This cube is then recursively subdivided into eight sub-cubes, until each cube contains at most one particle. The BBCP tree is constructed bottom up. Particles which are mutual nearest neighbours are replaced with a node and this process is repeated, using both particles and nodes until only one node is left. The forces are calculated in a similar way in both algorithms. In the BH code the forces are calculated relative to the node if

$$\frac{l}{d} < \theta$$

in the BBCP code if

$$\frac{R}{d} < \theta$$

where l and R are the size of the cube and the radius of the node respectively, d the distance from the particle to the node and θ the parameter which controls the accuracy as well as the calculation cost. If these criteria are not met the nodes are split into subnodes.

Makino [79] compared the Barnes and Hut and the Benz, Bowers, Cameron and Press algorithm. He concluded that for comparable accuracy the cost of the force calculation is similar for both methods. Tree construction takes longer with the BBCP scheme, but this is only a small part of the total calculations. The BH tree is easier to vectorize, though, [26] [35] which means it is considerably faster on vector processors. Steinmetz and Müller used a modification of the BBCP tree in order to deal with situations where there is more than one mutual nearest neighbour. Splinter et al. [122] performed a series of tests of agreement between PM, P^3M and tree codes. They concluded that PM codes are much faster than

the other two types. They conclude that the different methods agree well only on scales above the mean particle separation. This means, among other things, that the inner parts of dark matter halos in cosmological simulation cannot be resolved as they are typically below the mean interparticle separation.

The gravitational force needs to be softened in order to prevent the gravitational forces from going to infinity as the distance between the particles goes to zero. Softening also reduces 2 body relaxation effects. A simple way of softening the gravitational force is taking the potential as $\Phi \propto (r^2+\varepsilon^2)^{-1/2}$. An alternative way is by using a spline kernel (see also 4.3). The advantage of the latter method is that for separations larger than 2ε the potential is identical to the Kepler form [59]. The disadvantage is that it is more computationally expensive.

4.3 Smooth Particle Hydrodynamics (SPH)

4.3.1 Principles

Smooth particle hydrodynamics was introduced by Lucy [77] and further developed by Gingold and Monaghan [51]. The method is based on two principles:

1. The properties of the fluid at any point can be estimated by taking a weighted average of those properties over the surrounding volume.

2. The fluid can be represented by a finite number of particles.

For a continuum, the first point can be expressed mathematically as:

$$\hat{A}(\mathbf{r}) = \int A(\mathbf{r}')W(\mathbf{r}-\mathbf{r}',h)d^3r' \tag{4.2}$$

where $\hat{A}(\mathbf{r})\rangle$ is the estimated value of a physical quantity $A(\mathbf{r})$. $W(\mathbf{r},h)$ is an interpolating kernel and h, the smoothing length, represents the SPH smoothing

scale. $W(\mathbf{r}, h)$ has the following properties:

$$\int W(\mathbf{r} - \mathbf{r}', h) d^3 r' = 1 \tag{4.3}$$

and

$$\lim_{h \to 0} W(\mathbf{r} - \mathbf{r}', h) = \delta(\mathbf{r} - \mathbf{r}') \tag{4.4}$$

The second point states that the fluid can be represented by a finite number of particles. If, these particles are randomly distributed in such a way that for equal mass particles the probability of being found in the volume element dV at $\mathbf{r}$ is proportional to $\rho(\mathbf{r})\, dV$, equation 4.2 can be written as a Monte Carlo approximation:

$$\hat{A}(\mathbf{r}) = \sum_b m_b \frac{A_b}{\rho_b} W(\mathbf{r} - \mathbf{r}_b, h) \tag{4.5}$$

where A_b is the value of the quantity A, m_b and $\mathbf{r}_b$ are the mass and position for particle b, respectively, and ρ_b is the density at $\mathbf{r}_b$. The summation is over all particles. The error of approximating equation 4.2 by 4.5 is typically of $O(h^2)$. In particular the density is estimated everywhere by:

$$\rho(\mathbf{r}) = \sum_b m_b W(\mathbf{r} - \mathbf{r}_b, h) \tag{4.6}$$

There are two interpretations of equation 4.6 [59]. The first interpretation, the "scatter" interpretation, says that each particle has a mass which is smeared out in space. The density at any point in space is found by summing the contributions from the density profile of neighbouring particles. The second interpretation, the

"gather" interpretation, regards the particles as point markers in the fluid. Local properties at any point in space are obtained by sampling all neighbouring particles and weighting their contribution. The difference of the two interpretations is in the smoothing lengths used in equation 4.6. The "scatter" interpretation uses the smoothing length of particle b, whereas the "gather" interpretation uses the smoothing length of the particle at $\mathbf{r}$. If the smoothing length is the same for all particles, this distinction vanishes but it becomes relevant if h is spatially variable.

SPH provides a natural means to obtain gradients of local fluid properties. As long as the kernel $W(\mathbf{r} - \mathbf{r}')$ is symmetric in $\mathbf{r}$ and $\mathbf{r}'$, partial derivatives of physical quantities can be transformed into spatial derivatives of the kernel, using integration by parts. ∇A, for example, can be approximated by

$$\nabla A(\mathbf{r}) = \sum_b m_b \frac{A_b}{\rho_b} \nabla W(\mathbf{r} - \mathbf{r}_b, h) \tag{4.7}$$

A higher accuracy can be can be obtained by [84]

$$\rho \nabla A = \nabla(\rho A) - A \nabla \rho \tag{4.8}$$

It is now possible to find SPH equivalents of the flow equations. The flow equations in lagrangian form, ignoring viscosity and radiative cooling, are the continuity equation,

$$\frac{d\rho}{dt} = -\rho \nabla \cdot \mathbf{v} \tag{4.9}$$

where $\mathbf{v}$ is the velocity, the momentum equation,

$$\frac{d\mathbf{v}}{dt} = -\frac{1}{\rho}\nabla P - \nabla\phi \tag{4.10}$$

with P the pressure and ϕ the gravitational potential, and the energy equation

$$\frac{du}{dt} = -\frac{P}{\rho}\nabla \cdot \mathbf{v} \tag{4.11}$$

with u the specific internal energy. In SPH, if we regard each particle as representing a fixed mass of fluid then the continuity equation can be replaced by equation 4.6. Alternatively we can use the explicit

$$\frac{d\rho_a}{dt} = \sum_b m_b(\mathbf{v}_a - \mathbf{v}_b)\nabla_a W_{ab} \tag{4.12}$$

where $\nabla_a W_{ab}$ is the gradient of $W(\mathbf{r}_a - \mathbf{r}_b, h)$ with respects to the coordinates of particle a.

It is easiest to produce flow equations that preserve momentum and energy if the effective interparticle forces obey Newton's third law. This is achieved by taking a symmetrized form for the pressure gradient [84]

$$\frac{d\mathbf{v}_a}{dt} = -\sum_b m_b \left(\frac{P_b}{\rho_b^2} + \frac{P_a}{\rho_a^2}\right) \nabla_a W_{ab} \tag{4.13}$$

The energy equation can be written in a form which has the same symmetric factors

$$\frac{du_a}{dt} = \frac{1}{2}\sum_b m_b \left(\frac{P_b}{\rho_b^2} + \frac{P_a}{\rho_a^2}\right) \mathbf{v}_{ab} \cdot \nabla_a W_{ab} \tag{4.14}$$

where $\mathbf{v}_{ab} = \mathbf{v}_a - \mathbf{v}_b$. When artificial viscosity, which is necessary to capture shocks, is added, equation 4.13 and 4.14 are the most commonly used forms of

the momentum and energy equation.

4.3.2 Kernels and smoothing lengths

In the more than twenty years since the first appearance of SPH, a wide variety of implementations have been described in the literature. Apart from the flow equations, areas of difference include: the type of kernel, the smoothing lengths, the form of artificial viscosity used, the way time stepping is done and the physics included in the simulation.

The types of kernel used fall in two main categories, gaussian and super gaussian kernels and spline kernels. While a gaussian kernel is usually assumed to get a physical interpretation of SPH [84], in practice most researchers use a spline kernel. Although gaussian kernels

$$W(r,h) = \frac{1}{h^{\nu}\pi^{\nu/2}} e^{-(r/h)^2} \tag{4.15}$$

where ν is the number dimensions, interpolate with high accuracy, they have the disadvantage that their value does not go to zero at a finite distance. In order to avoid having an excessively large number of particles under the kernel the kernel has to be truncated at some distance, usually 3 or 4 smoothing lengths. The advantage of spline kernels is that they are zero outside a few smoothing lengths. The most often used spline kernel is the so called B_2-spline, proposed by Monaghan and Latanzio [86], which is in 3 dimensions

$$W(r,h) = \frac{W_s(r/h)}{h^3} \tag{4.16}$$

with

$$W_s(r/h) = \frac{1}{4\pi} \begin{cases} 4 - 6(r/h)^2 + 3(r/h)^3 & 0 \leq (r/h) \leq 1; \\ (2 - (r/h))^3, & 1 < (r/h) \leq 2; \\ 0, & (r/h) > 2. \end{cases} \quad (4.17)$$

This kernel has a continuous second derivative. The gradient of the kernel is sometimes modified to have a non-zero value for $r = 0$ [128]. The main reason for this is to create a small repulsive force for close particles to prevent them from clustering together. This makes the particles behave more like particles and less like fluid elements.

In the early implementations of SPH smoothing lengths were taken to be variable in time but constant in space. The smoothing lengths were calculated by

$$h \propto 1/\langle\rho\rangle^{1/\nu} \quad (4.18)$$

with $\langle\rho\rangle$ the average density of the particles and ν the number of dimensions. This is computationally intensive in high density areas and limits the resolution. Spatially variable smoothing length which are based on the local particle density [59] [43] greatly improved the resolution.

There are three basic ways of calculating the smoothing lengths. The first is similar to the method mentioned above but takes the local density instead of the average density

$$h_a \propto 1/\rho_a^{1/\nu} \quad (4.19)$$

The second method calculates the rate of change of h, using the relation above and the continuity equation

$$\frac{dh_a}{dt} = -\left(\frac{h_a}{\nu\rho_a}\right)\frac{d\rho_a}{dt} = \frac{1}{\nu}h_a\nabla\cdot\mathbf{v}_a \tag{4.20}$$

Both methods are fast. The second method may lead to a large accumulated error but it is still the preferred method of many authors (e.g. [95]). A problem with these methods is that they implicitly assume that the number density of the particles is proportional to the mass density. If this is not the case, the number of particles can vary considerably. The third method does not consider the density but tries to control the number of particles under the kernel directly. A popular way of doing this is the method described by Hernquist and Katz [59]

$$h_a^n = \frac{h_a^{n-1}}{2}\left[1 + \left(\frac{N_s}{N_a^{n-1}}\right)^{1/\nu}\right] \tag{4.21}$$

where h_a^n is the smoothing length at time-step n for particle a, N_s is the desired number of neighbours under the kernel and N_a^{n-1} is the number of particles under the kernel at time-step $n-1$. This method works well when the density gradients are small. It can, however, lead to oscillation in the estimates of h and the number of particles under the kernel if there are large density gradients.

While smoothing over a too large number of particles leads to a loss of resolution and is computationally expensive, it is less problematic than smoothing over too few particles. Smoothing over too few particles can lead to spurious shocking of the gas. Under unfavorable circumstances this can lead to a large increase in the energy content of the gas. Thacker et al. [127] describe an algorithm for updating the smoothing lengths which results in a fairly constant number of particles under the kernel. Instead of just counting all particles under the kernel when they calculate N_a^{n-1}, they limit the contribution of the outer

particles, using the kernel, and only count the particles within $\frac{3}{2}h$ at full weight. The smoothing lengths are updated using a weighting function which limits the relative change in the smoothing lengths.

A consequence of spatially variable smoothing lengths is that the "gather" and "scatter" interpretation of SPH give different results. Furthermore, since the effective interparticle force is not symmetric, in both cases linear and angular momentum are not conserved. This problem can be circumvented by symmetrizing the kernel. This can be done by taking the smoothing length of a particle as the average of the two smoothing lengths [43]. The density is then calculated by

$$\rho(\mathbf{r_a}) = \sum_b m_b W(\mathbf{r}_a - \mathbf{r}_a, h_{ab}) \tag{4.22}$$

where $h_{ab} = (h_a + h_b)/2$. It is also possible to use other averages as long as they are symmetric in a and b. Another method to achieve force symmetry is to take an average of the kernels [59]. In this case the density is given by

$$\rho(\mathbf{r}) = \sum_b m_b[W(\mathbf{r}_a - \mathbf{r}_b, h_a) + W(\mathbf{r}_a - \mathbf{r}_b, h_b)]/2 \tag{4.23}$$

and the momentum equation becomes

$$\frac{d\mathbf{v}_a}{dt} = -\sum_b m_b \left(\frac{P_b}{\rho_b^2} + \frac{P_a}{\rho_a^2}\right) \times \nabla_a W(\mathbf{r}_a - \mathbf{r}_b, h_b)]/2 \tag{4.24}$$

Thomas and Couchman [128] use yet another method, calculating the density according to the scatter interpretation and symmetrizing the momentum equations as

$$\frac{d\mathbf{v}_a}{dt} = -\sum_b m_b \frac{P_a}{\rho_a^2} \nabla_a W(\mathbf{r}_a - \mathbf{r}_b, h_a) + \sum_b m_b \frac{P_b}{\rho_b^2} \nabla_b W(\mathbf{r}_a - \mathbf{r}_b, h_b) \qquad (4.25)$$

This is the form of the momentum equation used in the publicly available code, HYDRA [34].

Variable smoothing lengths result in additional, ∇h terms in the equations of motion. Usually these terms are neglected, because they are small compared to the other terms, and because calculating them would be computationally expensive. However, Hernquist [60] has shown that under unfavorable circumstances errors in global quantities can amount to $\sim$ 10% and that in simulations of adiabatic systems simultaneous conservation of energy and entropy cannot be obtained. Nelson and Papaloizou [96] and Serna et al. [113] showed that the inclusion of the ∇h terms results in an dramatic improvement in accuracy for problems in which SPH normally does not conserve energy or entropy well. To prevent the extra computations which result from including these terms while maintaining accuracy, Capuzzo-Dolcetta and Di Lisio [24] proposed a criterion to select kernels in which the errors caused by neglecting the ∇h term would be minimized. They applied this criterion to the category of spline kernels and found that the preferred kernel is not the almost universally applied B_2 spline of Monaghan and Latanzio [86].

Anisotropic kernels where the smoothing lengths are represented by a tensor instead of by a scalar have also been applied [114] [102]. This method, called Adaptive Smooth Particle Hydrodynamics (ASPH), results in a better resolution in situations with strong anisotropic volume changes. An important shortcoming of the method, however, is that angular momentum is not conserved. For a

comparison of ASPH with other implementations of SPH and with Eulerian codes see [49].

4.3.3 Artificial Viscosity

Another area where the various implementation of SPH differ is the form of artificial viscosity which is used. Artificial viscosity is necessary to prevent interpenetration of particles in shocks. The most common form of artificial viscosity in SPH simulations is the "standard" Monaghan and Gingold [85] form

$$\Pi_{ab} = \begin{cases} \frac{-\alpha c_{ab}\mu_{ab}+\beta\mu_{ab}^2}{\rho_{ab}} & \mathbf{v}_{ab}\cdot\mathbf{r}_{ab} < 0; \\ 0, & \mathbf{v}_{ab}\cdot\mathbf{r}_{ab} > 0; \end{cases} \tag{4.26}$$

with

$$\mu_{ab} = \frac{h\mathbf{v}_{ab}\cdot\mathbf{r}_{ab}}{\mathbf{r}_{ab}^2+\eta^2} \tag{4.27}$$

where $\rho_{ab} = (\rho_a + \rho_b)/2$ is the average density and $c_{ab} = (c_a + c_b)/2$ the average sound speed of a pair of particles, $\mathbf{v}_{ab} = \mathbf{v}_a - \mathbf{v}_b$ and $\mathbf{r}_{ab} = \mathbf{r}_a - \mathbf{r}_b$. $\eta^2 = 0.01h_{ij}^2$ is used to avoid numerical divergence. α and β are constants of order unity. The first term on the RHS of equation 4.26 produces shear or bulk viscosity, the second term is essentially the standard Von Neumann-Richtmyer artificial viscosity.

Balsara [6] added a shear correction term to this equation. In the presence of shear flows $\nabla\cdot\mathbf{v} = 0$, $|\nabla\times\mathbf{v}| > 0$, the standard form of the artificial viscosity leads to shear viscosity and therefore to angular momentum transport. To prevent this Π_{ab} is multiplied by a term

$$f_{ab} = \frac{f_a + f_b}{2} \tag{4.28}$$

with

$$f_a = \frac{|\langle \nabla \cdot \mathbf{v} \rangle_a|}{|\langle \nabla \cdot \mathbf{v} \rangle_a| + |\langle \nabla \times \mathbf{v} \rangle_a| + 0.0001 c_a / h_a} \tag{4.29}$$

This form of the artificial viscosity was implemented by Steinmetz [124] and Navarro and Steinmetz [94], who found that the dissipation of angular momentum is drastically reduced in problems with small numbers of particles. Of course it is also possible that shear viscosity is desired in simulations. One way of adding shear viscosity is by taking the standard artificial viscosity and taking $\beta = 0$ [89]. In simulations of galaxy formation shear viscosity, beyond the normal artificial viscosity, has seldom been included [78].

An alternative form of the artificial viscosity is given by Thomas and Couchman [128]

$$P_i \rightarrow \begin{cases} P_i + \rho_i[-\alpha c_i h_i \nabla \cdot v_i + \beta (h_i \nabla \cdot v_i)^2], & \nabla \cdot v_i < 0; \\ P_i, & \nabla \cdot v_i \geq 0, \end{cases} \tag{4.30}$$

Thacker et al. [127] compared implementations of SPH using different ways of symmetrizing the momentum equation combined with different forms of artificial viscosity in 7 test scenarios. They concluded that the artificial viscosity is the single most important factor distinguishing the results from various implementations. The artificial viscosity proposed by Thomas and Couchman [128] performs rather poorly and the standard artificial viscosity, if necessary with the correction for the shear viscosity, is a better choice. The way the force symmetrization

is achieved is less important, although most of their results favour kernel symmetrization. There are indications that the ideal values of the coefficient α and β depend on the situation in which it is applied [59] [64], with stronger shocks requiring larger values of the coefficients. Lombardi et al. [76] performed a number of tests investigating the effects of spurious transport in 3-dimensional SPH. They found that if it is important to control spurious mixing, a strong artificial viscosity is favorable. For most situations they recommend $\alpha = 0.5$, $\beta = 1$.

4.3.4 Time-stepping

The time integration is usually performed using either a simple leapfrog scheme [52] [59] or a second order Runge-Kutta method [95]. In order to obtain second order accuracy both methods require evaluation of the forces at half time. The leapfrog integrator calculates positions as

$$\mathbf{r}_a^{n+1/2} = \mathbf{r}_a^{n-1/2} + \Delta t\, \mathbf{v}_a^n \tag{4.31}$$

The velocities are updated in two stages. First a predicted estimate is obtained from

$$\mathbf{v}_a^{n+1/2} = \mathbf{v}_a^n + \frac{\Delta t}{2}\dot{\mathbf{v}}_a^{n-1/2} \tag{4.32}$$

and this is used to compute the acceleration at half time-step $\dot{\mathbf{v}}_a^{n+1/2}$. The velocity is then updated as

$$\mathbf{v}_a^{n+1} = \mathbf{v}_a^n + \Delta t\, \dot{\mathbf{v}}_a^{n+1/2} \tag{4.33}$$

The energy is updated as

$$u_a^{n+1/2} = u_a^{n-1/2} + \frac{\Delta t}{2}(\dot{u}_a^{n+1/2} + \dot{u}_a^{n-1/2}) \tag{4.34}$$

In the Runge-Kutta method position, velocity and energy are updated as

$$\mathbf{r}_a^{n+1} = \mathbf{r}_a^n + f\mathbf{v}_a^{n+1/2}\Delta t + (1-f)\mathbf{v}_a^n \Delta t \tag{4.35}$$

$$\mathbf{v}_a^{n+1} = \mathbf{v}_a^n + f\dot{\mathbf{v}}_a^{n+1/2}\Delta t + (1-f)\dot{\mathbf{v}}_a^n \Delta t \tag{4.36}$$

$$u_a^{n+1} = u_a^n + f\dot{u}_a^{n+1/2}\Delta t + (1-f)\dot{u}_a^n \Delta t \tag{4.37}$$

where f is a constant that depends on the particular implementation of the Runge-Kutta method.

There are many different ways to set the time-step [86] [84] [95] [127] [131]. Most implementations determine the time-step based on the relative velocities of the particle and the kernel length, e.g.

$$\delta t = C \min_a (h_a / max_b(|\mathbf{v}_a - \mathbf{v}_b|) \tag{4.38}$$

where C is a factor of order unity, which typically has a value of $0.1 - 0.3$. This condition is inadequate when relative particle motions are small and it is necessary to allow for changes due to the pressure forces. In that case the time-step needs to depend on the sound speed as well. Other things which can be taken into account are the acceleration, the cooling time (if cooling is included in the simulation) and the artificial viscosity. One often used time-step is

$$\delta t = C \min(\delta t_f, \delta t_s) \tag{4.39}$$

with C the Courant factor, which typically has a value of $0.1 - 0.3$ and

$$\delta t_f = \min_a \left(\frac{h_a}{|\mathbf{f}_a|} \right)^{1/2} \tag{4.40}$$

where $\mathbf{f}_a$ the acceleration, and

$$\delta t_s = \min_a \frac{h}{c_a + 0.6(\alpha c_a + \beta \max_b \mu_{ab})} \tag{4.41}$$

where c_a is the sound speed of particle a, μ_{ab} the artificial viscosity parameter, given by equation 4.27 and α and β the artificial viscosity parameters. There have been no systematic studies to evaluate the different methods used to determine the time-step. Multiple time-steps, where different particles have different time-step are often used in simulations with large numbers of particles (e.g. [59] [26] [35]). In such cases a scheme is used where the time-steps can only vary by factors of two and where the times are regularly synchronized.

4.3.5 Radiative cooling and star formation

Apart from the differences in the numerical schemes used, implementations of SPH also differ in the physics that is included. For simulations of galaxy formation the most relevant processes not mentioned yet are radiative cooling and star formation. Cosmological simulations with cooling [69] [92] [48] [44] broadly confirm the standard view of galaxy formation as described by White and Rees [139]. The gas in galaxies is much more concentrated than the dark matter and

the visible parts of galaxies survive the mergers of their halos in clusters of galaxies. However, so far simulations have not been able to produce disc sizes similar to observed spirals. In these simulations a large proportions of the gas, $\sim$ 70%, is located in a central knot and the specific angular momentum of the gas is too small. Usually it is assumed that star formation and the associated feedback from supernovae will help alleviate this problem [41] [91] [124] [120].

Another problem is a deficit of hot gas produced in the collapse due to in-shock cooling [78]. Limited spatial resolution means that the gas spends much more time traveling through simulated shocks than it does in reality. This means it can cool significantly during the shocks and fails to reach the correct post-shock temperature. A more detailed discussion of this problem will be given in chapter 6. Another cause of resolution related excessive cooling was described by Thacker et al. [127]. If dense knots of cold gas are surrounded by hot gas with a much lower density then it is possible that the smoothing radius of a particle in the hot surrounding gas encompasses the cold clump. This results in smoothing over an excessive number of cold particles leading to an overestimation of the density of the hot particle. Since radiative cooling is proportional to the density, the cooling of the hot gas particle is overestimated. Keeping the number of particles under the kernel within narrow limits will alleviate this problem.

Lack of resolution may also result in insufficient cooling or excessive heating. In cosmological simulations, individual galaxies are often represented by a very small number of particles. This may set an artificial upper limit on the density, resulting in insufficient cooling [69]. It also results in a spurious transport of energy from the dark matter to the gas particles due to two body relaxation

effects [126]. Using a larger number of dark matter than gas particles may reduce this problem.

Several authors have included star formation [67] [68] [82] [95] [133]. Incorporating star formation requires a determination of the star formation rate and a decision on how to implement this. Navarro and White [95], and Steinmetz and Müller [125] formed new star particles from the gas particles, reducing the mass of the gas particles accordingly. A disadvantage of this method is that it may lead to an excessive number of particles, significantly slowing calculations. An alternative way of incorporating star formation is to convert entire gas particles into star particles. This is a crude method, however, which does not allow for a gradual conversion of gas into stars. Mihos and Hernquist [82] used a method in which each gas particle has a total mass and a gas mass. As stars form, the gas mass is reduced while the total mass remains fixed. In other words the particles are hybrid gas-star particles. A disadvantage of this method is that gas and stars remain kinematically coupled, which means that the stars will move under the influence hydrodynamic effects.

An important consequence of star formation is supernova feedback. Since star formation and supernova feedback are processes which are both poorly understood and unresolvable in SPH simulation of galaxy formation (a typical SPH particle is many orders of magnitude larger than a star) they have to be included in a heuristic way. Katz [67] included the feedback from supernovae by reducing the mass of their star particles and adding heat to the surrounding gas particles. Metzler and Evrard [81] included the effect of ejection of gas from galaxies by supernova explosions in their simulations of a cluster of galaxies. Their simulations

did not involve the creation of stars but did take the effect of supernova feedback into account. Navarro and White [95] modeled the supernova feedback by adjusting both the thermal and the kinetic energy of the surrounding gas particles. They did this by giving the neighbouring gas particles a velocity perturbation $\Delta \mathbf{v}$ radially away from the star particle. Mihos and Hernquist [82] used a similar method but instead of giving all neighbours the same velocity perturbations they used the kernel to distribute the kinetic energy among the gas particles, giving close gas particles larger velocity perturbations than the more distant particles. The result of simulations including supernova feedback is that, since supernovae release their energy in high density regions, the resulting thermal energy is radiated quickly. Kinetic energy does have an influence and reduces the density and the amount of star formation but this is highly dependent on the method used to model supernova feedback [95]. Mori et al. [87] simulated the formation of a dwarf galaxy, incorporating the effects of supernova driven winds, and concluded that this process was likely to be the decisive mechanism in the formation of dwarf galaxies.

Chapter 5

The 2D and 3D SPH codes

5.1 Introduction

Two galaxy formation models are used for the second part of this research. Both are N-body models combined with Smooth Particle Hydrodynamics (SPH) and both simulate the formation of a single disc galaxy. The models contain a dark matter component, represented by collisionless dark matter particles, and a gas component, modeled with SPH. As explained in the previous chapter, the two main principles of SPH are that a fluid or gas can be represented by a finite number of particles and that the properties of the fluid at each point can be determined by taking a weighted average of the properties of the neighbouring particles. The gas is therefore represented by particles, carrying properties like mass and temperature. While the dark matter particles only move under the influence of gravity, the motion of the gas particles is determined by gravitation, pressure and artificial viscosity.

The main difference between the two models is in the number of dimensions. The first model is a 2-dimensional (2D), axisymmetric model while the second model is a more conventional 3-dimensional (3D) model. The advantage of using a lower number of dimensions is that, with a similar number of particles, a much higher resolution can be achieved. For example, 10^4 particles in 2 dimensions result in the same resolution as 10^6 particles in 3 dimensions. As mentioned in the previous chapter, the most time consuming part of N-body methods is the gravitational force calculation. Direct summation of the forces between pairs of particles would lead to $O(N^2)$ calculations, but this can be reduced to $O(NlogN)$ by applying techniques like treecodes. This means that a simulation in 3 dimensions takes $\sim \frac{3}{2}\sqrt{N}$ as long as a simulation with the same resolution in 2 dimension,

where N is the number of particles used in 2 dimensions. For the simulations done here with the 2D code, this is about a factor 50.

The main disadvantage of a 2D code is that all information in one dimension is lost. In particular it is not possible to form spiral structures with an axisymmetric model. Another disadvantage is that the force calculations are more complicated because the particles represent rings of matter instead of point particles. To some extent this can be overcome by using look-up tables. Some other disadvantages will be discussed later.

The 2-dimensional code was written by Maguire and is described in detail in his Ph.D. thesis [78]. Although his code was completely rewritten for this work, its essential elements are unaltered. The 3-dimensional model was written by the author. Originally the model was intended as a 3-dimensional version of Maguire's code, which meant that parts of the code which did not depend on the dimension were the same as in Maguire's code. However, some changes were made later. In particular the kernel was changed to the B_2 spline and the method of Thacker et al. [127] for updating the smoothing lengths was adopted. In the simulations presented here, the later version of the model is used.

5.2 Gravitational Force Calculation

In the 3D galaxy formation model particles are represented by point particles. The gravitational potential due to a particle is

$$\phi(\mathbf{r}) = -\frac{Gm_0}{|\mathbf{r} - \mathbf{r}_0|} \tag{5.1}$$

with G the gravitational constant and m_0 and $\mathbf{r}_0$ the mass and position of the

particle respectively. The acceleration is

$$\mathbf{a} = -\frac{Gm_0}{|\mathbf{r} - \mathbf{r}_0|^3}(\mathbf{r} - \mathbf{r}_0) \tag{5.2}$$

The gravitational potential is softened, using a fixed softening length, $\varepsilon = 0.2 \times 10^{-2}$ in units of the code.

In the 2D galaxy formation model each particle represents a ring of mass. The particles have different masses with the mass of the particles proportional to their cylindrical radius at the start of the simulation. The potential due to a ring of mass m_0 with cylindrical radius ϖ_0 at z_0 is:

$$\phi(\varpi, z) = -\frac{2Gm_0}{\pi b^{1/2}} K(a) \tag{5.3}$$

where

$$b = (\varpi + \varpi_0)^2 + (z - z_0)^2, \qquad a = \frac{4\varpi\varpi_0}{b} \tag{5.4}$$

and $K(a)$ is an elliptical integral [1]

$$K(a) = \int_0^{\pi/2} \frac{d\theta}{(1 - a\sin^2\theta)^{1/2}} \tag{5.5}$$

From the potential, the acceleration can be derived. The component in the ϖ direction is

$$a_\varpi = -\frac{2Gm_0}{\pi}\left[-\frac{(\varpi + \varpi_0)K(a)}{b^{3/2}} + \frac{1}{2a}\frac{E(a)}{1-a} - K(a)\frac{1}{b^{5/2}}4\varpi_0 b - 2\varpi(\varpi + \varpi_0)\right] \tag{5.6}$$

with a and b as in 5.4 and $K(a)$ as in 5.5. $E(a)$ is also an elliptical integral [1]

$$E(a) = \int_0^{\pi/2} \sqrt{1 - a\sin^2\theta)}d\theta \tag{5.7}$$

The component of the acceleration due to gravity in the z direction is

$$a_z = -\frac{2Gm_0}{\pi}\left[-\frac{(z-z_0)K(a)}{b^{3/2}} - \frac{1}{2a}[\frac{E(a)}{1-a} - K(a)]\frac{1}{b^{5/2}}8\varpi_0\varpi(z-z_0)\right] \tag{5.8}$$

Self gravity of the rings is ignored. The gravitational force is softened with a softening length of 2×10^{-2} in the distance units of the code.

5.3 Tree construction

Gravitational forces are calculated using the tree construction method of Benz et al. [12], which was described in section 4.2.

In the axisymmetric model the nodes represent rings of mass, just as the particles do. The mass of a node is the combined mass of the particles that are part of that node. The location of the node is a more difficult issue. The z coordinate is the z coordinate of the centre of mass of the children of the node

$$z = \frac{m_1 z_1 + m_2 z_2}{M} \tag{5.9}$$

with m_1, m_2 and z_1, z_2 the masses and z coordinates of the children of the node and M the mass of the node. The ϖ coordinate is calculated in a similar way

$$\varpi = \frac{m_1 \varpi_1 + m_2 \varpi_2}{M} \tag{5.10}$$

Although this is not necessarily the best approach, it was found that errors in the axisymmetric force calculation are comparable to the cartesian case [78].

The radius of a node is calculated to ensure that all particles which are children of the node are closer to the node than this radius. In the axisymmetric case this can be seen as the thickness of the ring. The radius is calculated as in BBCP:

$$R = \max(\frac{m_1}{M}|\mathbf{r_2} - \mathbf{r_1}| + R_2, \frac{m_2}{M}|\mathbf{r_2} - \mathbf{r_1}| + R_1) \tag{5.11}$$

where m_1, m_2, $\mathbf{r_2}, \mathbf{r_1}$ and R_1, R_2 are the masses, positions and radii of the subnodes respectively.

A node opening criterion determines if the gravitational forces are calculated with respect to a node or with respect to the children of that node. Forces are calculated relative to the node if

$$\frac{R}{d} < \theta \tag{5.12}$$

with R the radius of the node, d the distance from the particle to the node and θ the opening parameter, which has the value 0.6 for these simulations. If this criterion is not met the nodes are split into subnodes.

5.4 Flow equations

The flow equations in the lagrangian form, ignoring viscosity and radiative cooling, are the continuity equation,

$$\frac{d\rho}{dt} = -\rho\nabla \cdot \mathbf{v} \tag{5.13}$$

the momentum equation,

$$\frac{d\mathbf{v}}{dt} = -\frac{1}{\rho}\nabla P - \nabla\phi \tag{5.14}$$

and the energy equation

$$\frac{du}{dt} = -\frac{P}{\rho}\nabla\cdot\mathbf{v} \tag{5.15}$$

For an axisymmetric system these can be written in cylindrical coordinates as

$$\frac{d\rho}{dt} = -\rho\left[\frac{1}{\varpi}\frac{\partial}{\partial\varpi}(\varpi v_\varpi) + \frac{\partial v_z}{\partial z}\right] \tag{5.16}$$

$$\frac{dv_\varpi}{dt} = -\frac{1}{\rho}\frac{\partial P}{\partial\varpi} - \frac{\partial\phi}{\partial\varpi} + \frac{l_z^2}{\varpi^3} \tag{5.17}$$

$$\frac{dv_z}{dt} = -\frac{1}{\rho}\frac{\partial P}{\partial z} - \frac{\partial\phi}{\partial z} \tag{5.18}$$

$$\frac{du}{dt} = -\frac{P}{\rho}\left[\frac{1}{\varpi}\frac{\partial}{\partial\varpi}(\varpi v_\varpi) + \frac{\partial v_z}{\partial z}\right] \tag{5.19}$$

where v_ϖ and v_z are the components of the velocity in the ϖ and z direction, respectively, and l_z is the angular momentum of the rings around the z axis. In SPH the density is calculated as

$$\rho(\mathrm{r_i}) = \sum_j m_j W_{ij} \tag{5.20}$$

in both the 2D and the 3D model. The momentum and the energy equation for the 3D model, using Monaghan's [84] symmetrisation of the pressure force, are written as

$$\frac{d\mathbf{v}_i}{dt} = -\sum_j m_j \left(\frac{P_i}{\rho_i^2} + \frac{P_j}{\rho_j^2}\right) \nabla_i W_{ij} - \nabla\phi \tag{5.21}$$

$$\frac{du_i}{dt} = \frac{1}{2}\sum_j m_j \left(\frac{P_i}{\rho_i^2} + \frac{P_j}{\rho_j^2}\right) \mathrm{v}_{ij} \cdot \nabla_i W_{ij} \tag{5.22}$$

In the 2 dimensional case, these are

$$\frac{dv_{\varpi,i}}{dt} = \sum_j m_j \left(\frac{P_i}{\rho_i^2} + \frac{P_j}{\rho_j^2}\right) \frac{\partial W_{ij}}{\partial \varpi_i} - \frac{\partial \phi}{\partial \varpi_i} + \frac{l_{z,i}^2}{\varpi_i^3} \tag{5.23}$$

$$\frac{dv_{z,i}}{dt} = \sum_j m_j \left(\frac{P_i}{\rho_i^2} + \frac{P_j}{\rho_j^2}\right) \frac{\partial W_{ij}}{\partial z_i} - \frac{\partial \phi}{\partial z_i} \tag{5.24}$$

$$\begin{aligned} \frac{du_i}{dt} &= \frac{1}{2}\sum_j m_j \left(\frac{P_i}{\rho_i^2} + \frac{P_j}{\rho_j^2}\right) \\ &\times \left[\left(\frac{\varpi_j}{\varpi_i} v_{\varpi,j} - v_{\varpi,i}\right) \frac{\partial W_{ij}}{\partial \varpi_i} + (v_{z,j} - v_{z,i}) \frac{\partial W_{ij}}{\partial z_i}\right] \end{aligned} \tag{5.25}$$

In the axisymmetric code forces are calculated between rings. Since the rings represent a distribution of particles it is necessary to include a self-force terms. The self-force term of a ring can be seen as the force that arises from the fact that the particles which are part of the ring interact via their pressure forces. This force is negligible away from the symmetry axis, but it becomes appreciable when the "particles" of a ring interact strongly near the axis. To be consistent a self-gravity term should also be included. The self-gravity of a thin ring goes to infinity but it is possible to include a self-gravity term by calculating the self-gravity that would arise if the ring consisted of a fixed number of particles at equal distances from each other. This was tried, with the number of particles

taken to be proportional to the mass of the ring. This number was chosen in such a way that the resolution in the ϕ direction is similar to the resolution in the z and ϖ direction. It was shown that including softening complicated the calculations, thereby reducing the speed of the code and that the self-gravity term was small relative to the other gravity terms. Self-gravity was therefore ignored for the simulations presented here.

5.5 Artificial viscosity

The artificial viscosity used here is the standard "Monaghan" artificial viscosity [85].

$$\Pi_{ij} = \begin{cases} \frac{-\alpha c_{ij}\mu_{ij} + \beta\mu_{ij}^2}{\rho_{ij}} & \text{if } \mathbf{v}_{ij} \cdot \mathbf{r}_{ij} < 0; \\ 0, & \text{otherwise} \end{cases} \tag{5.26}$$

$$\mu_{ij} = \frac{h_{ij}\mathbf{v}_{ij} \cdot \mathbf{r}_{ij}}{r_{ij}^2 + \eta^2} \tag{5.27}$$

with $\rho_{ij} = (\rho_i + \rho_j)/2$, the average density and $c_{ij} = (c_i + c_j)/2$ the average sound speed of a pair of particles. $\mathbf{v}_{ij} = \mathbf{v}_i - \mathbf{v}_j$ and $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$. $\eta^2 = 0.01h_{ij}^2$ is used to avoid numerical divergence. $\alpha = 0.5$ and $\beta = 1$, as recommended by Lombardi et al. [76], although it should be noted that the best value for these constants is likely to be situation dependent. The first term on the RHS of equation 5.26 produces shear and bulk viscosity, the second term is essentially the standard Von Neumann-Richtmyer artificial viscosity.

Adding the artificial viscosity to the flow equations gives

$$\frac{d\mathbf{v}_i}{dt} = -\sum_j m_j \left(\frac{P_i}{\rho_i^2} + \frac{P_j}{\rho_j^2} + \Pi_{ij} \right) \nabla_i W_{ij} - \nabla\phi \tag{5.28}$$

$$\frac{du_i}{dt} = \frac{1}{2} \sum_j m_j \left(\frac{P_i}{\rho_i^2} + \frac{P_j}{\rho_j^2} + \Pi_{ij} \right) \mathrm{v}_{ij} \cdot \nabla_i W_{ij} \tag{5.29}$$

for the 3D model and

$$\frac{dv_{\varpi,i}}{dt} = \sum_j m_j \left(\frac{P_i}{\rho_i^2} + \frac{P_j}{\rho_j^2} + \Pi_{ij} \right) \frac{\partial W_{ij}}{\partial \varpi_i} - \frac{\partial \phi}{\partial \varpi_i} + \frac{l_{z,i}^2}{\varpi_i^3} \tag{5.30}$$

$$\frac{dv_{z,i}}{dt} = \sum_j m_j \left(\frac{P_i}{\rho_i^2} + \frac{P_j}{\rho_j^2} + \Pi_{ij} \right) \frac{\partial W_{ij}}{\partial z_i} - \frac{\partial \phi}{\partial z_i} \tag{5.31}$$

$$\begin{aligned} \frac{du_i}{dt} &= \frac{1}{2} \sum_j m_j \left(\frac{P_i}{\rho_i^2} + \frac{P_j}{\rho_j^2} + \Pi_{ij} \right) \\ &\quad \times \left[\left(\frac{\varpi_j}{\varpi_i} v_{\varpi,j} - v_{\varpi,i} \right) \frac{\partial W_{ij}}{\partial \varpi_i} + (v_{z,j} - v_{z,i}) \frac{\partial W_{ij}}{\partial z_i} \right] \end{aligned} \tag{5.32}$$

for the 2D model.

5.6 Radiative cooling

When radiative cooling is included an extra term, $-n_e n_h \Lambda/\rho$, where n_e and n_h are the electron and proton density, respectively, and Λ is the cooling function, must be added to the energy equations (equation 5.15 and 5.19). This is included in the model as $-\rho k \Lambda$, with k a conversion factor. The SPH expressions for the energy equation now become

$$\frac{du_i}{dt} = \frac{1}{2} \sum_j m_j \left(\frac{P_i}{\rho_i^2} + \frac{P_j}{\rho_j^2} + \Pi_{ij} \right) \mathrm{v}_{ij} \cdot \nabla_i W_{ij} - \rho k \Lambda \tag{5.33}$$

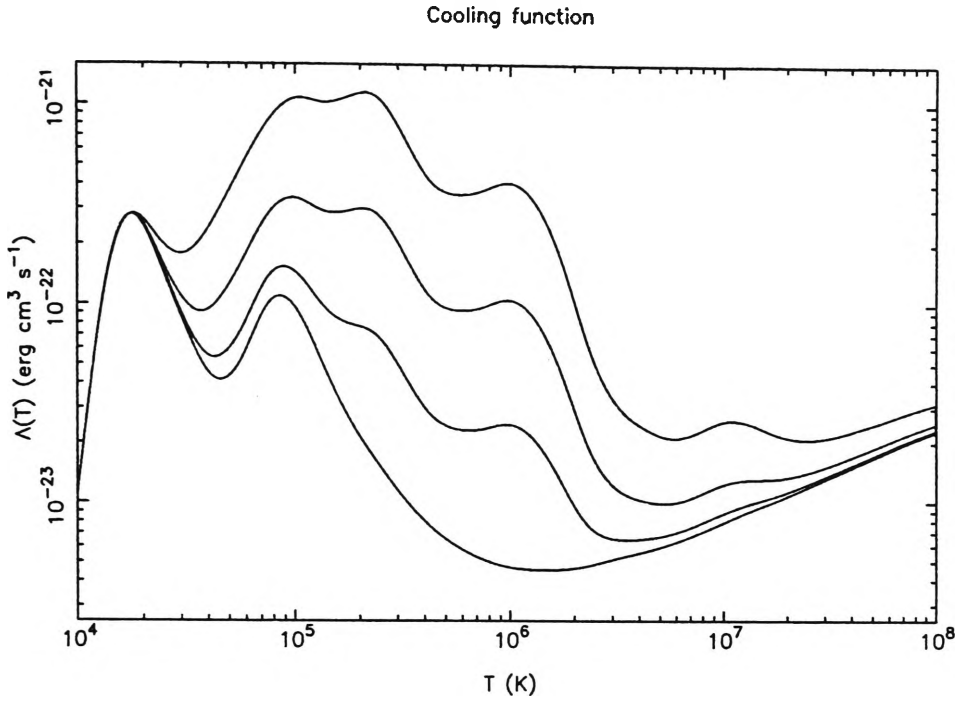

Figure 5.1:
Cooling function based on data from Böhringer and Hensler. From top to bottom, lines are for abudances (relative to solar) $Z/Z_\odot = 2$, 0.5, 0.1 and 0 respectively

for the 3D model and

$$\frac{du_i}{dt} = \frac{1}{2}\sum_j m_j \left(\frac{P_i}{\rho_i^2} + \frac{P_j}{\rho_j^2} + \Pi_{ij}\right) \times \left[\left(\frac{\varpi_j}{\varpi_i} v_{\varpi,j} - v_{\varpi,i}\right)\frac{\partial W_{ij}}{\partial \varpi_i} + (v_{z,j} - v_{z,i})\frac{\partial W_{ij}}{\partial z_i}\right] - \rho k \Lambda \qquad (5.34)$$

for the 2D model. The cooling function Λ is the same as the cooling function used for the semi-analytical models, based on the data from Böhringer and Hensler [17]. The value of the cooling function depends on the temperature and the abundance. A plot of the cooling function for several abundances is show in Figure 5.1.

5.7 The smoothing kernel

In an axisymmetric implementation of SPH the three dimensional kernel is replaced by a two dimensional kernel. This kernel is found by averaging over the cylindrical coordinate ϕ

$$\widetilde{W}(\varpi, \varpi', z, z') = \frac{1}{2\pi}\int_0^{2\pi} W(\mathbf{x} - \mathbf{x}')d\phi' \tag{5.35}$$

The value of a function can now be found as in conventional 3D SPH by

$$f(x) \simeq \sum_i \frac{f_i}{\rho_i} m_i \widetilde{W}(\varpi, \varpi_i', z, z_i) \tag{5.36}$$

If the 3-dimensional kernel, W, is a gaussian with a radius dependence of $\exp(r^2/2h^2)$, then

$$\begin{aligned}\widetilde{W}(\varpi, \varpi_i, z, z_i) &= \frac{1}{(2\pi)^{3/2}h^3} I_0\left(\frac{\varpi\varpi_i}{h^2}\right)\exp\left(-\frac{\varpi\varpi_i}{h^2}\right) \\ &\quad \exp\left(-\frac{1}{2h^2}[(\varpi - \varpi)^2 + (z - z_i)^2]\right)\end{aligned} \tag{5.37}$$

where I_0 is a modified Bessel function [1].

$$I_0(x) = \frac{1}{\pi}\int_0^{\pi} e^{x\cos\theta}d\theta \tag{5.38}$$

Since the gaussian kernel does not vanish at a finite distance, the kernel must be truncated. This is done at $4h$, where h is the smoothing length. The kernel is symmetric in ϖ and ϖ_i. An important feature of this kernel is that $\partial\widetilde{W}/\partial\varpi \neq 0$ when $\varpi = \varpi_i$. This gives rise to the self-force term as discussed in section 5.4, if the j=i term is included in sums for the force.

The 3 dimensional code was initially written with a Gaussian kernel which was also truncated at $4h$. Later the kernel was replaced by the B_2 spline kernel of Monaghan and Latanzio [86]

$$W(r,h) = \frac{W_s(r/h)}{h^3} \tag{5.39}$$

where W_s is

$$W_s(x) = \frac{1}{4\pi} \begin{cases} 4 - 6x^2 + 3x^3 & 0 \leq x \leq 1; \\ (2-x)^3, & 1 < x \leq 2; \\ 0, & x > 2. \end{cases} \tag{5.40}$$

where $x = (r/h)$. The derivative of the kernel is modified in order to give a small repulsive force for close particles [128]

$$\frac{dW_s(x)}{dx} = -\frac{1}{4\pi} \begin{cases} 4 & 0 \leq x \leq 2/3; \\ 3x(4-3x) & 2/3 < x \leq 1; \\ 3(2-x)^2, & 1 < x \leq 2; \\ 0, & x > 2. \end{cases} \tag{5.41}$$

The smoothing lengths vary in time and space. The kernel therefore needs to be symmetrized in order conserve momentum. This is done by taking the average of the smoothing length $h_{ij} = (h_i + h_j)/2$ and using this in W_{ij}. No attempt was made to include the ∇h correction terms.

5.8 Updating of smoothing lengths

Smoothing lengths in the 3D model were updated according to the method described in Thacker et al. [127]. This method tries to keep the weighted number of particles under the kernel constant. The weighted number of particles is calculated by

$$N_i = \sum_j W_{nn}(r_{ij}/h_{ij}) \tag{5.42}$$

with N_i the weighted number of particles under the kernel for particle i and

$$W_{nn}(r/h) = \begin{cases} 1 & 0 \leq r/h < 3/2; \\ \pi W_s(4(r/h - 3/2)), & 3/2 < r/h \leq 2; \end{cases} \tag{5.43}$$

with W_s as in equation 5.40. The smoothing lengths are updated using a weighting function which limits the relative change in h

$$h_i^n = h_i^{n-1}\left[1 - a + a\left(\frac{N_s}{N_i^{n-1}}\right)^{1/3}\right] \tag{5.44}$$

where N_s the desired weighted number of particles under the kernel, N_i^{n-1} the weighted number of particles under the kernel at step $n-1$ and a a weighting variable which is determined by

$$a = \begin{cases} 0.2(1+s^2) & s < 1; \\ 0.2(1+1/s^3), & s \geq 1; \end{cases} \tag{5.45}$$

with $s = (N_s/N_i^{n-1})^{1/3}$. The result is a scheme which has the asymptotic property $h_i^n = 0.8h_i^{n-1}$ but which yields $h_i^n = h_i^{n-1}[0.6+0.4(N_s/N_i^{n-1})^{1/3}]$ when the number of particles under the kernel is close to its target value.

Smoothing lengths in the 2D code were updated in a different way. Methods which try to keep the number of particles under the kernel constant sometimes cause problems if a particle nears the edge of the particle distribution. For three dimensional simulations this is not a big problem as the edges are on the outside of the particle distribution and the particles on the outside of the distribution do not have a major influence on the results of the simulation. For the two

dimensional simulations, however, there is also an edge on the symmetry axis at $\varpi = 0$. If a particle gets close to the axis the number of neighbouring particles can drop rapidly, resulting in large unphysical variations of the pressure force.

The smoothing lengths in the 2D code are integrated as

$$\frac{dh_i}{dt} = -\frac{1}{2}\frac{h_i}{\sigma_i}\frac{d\sigma_i}{dt} = \frac{1}{2}h_i\left(\frac{\partial v_\varpi}{\partial \varpi} + \frac{\partial v_z}{\partial z}\right) \tag{5.46}$$

where σ is the 2-dimensional number density of particles in the $\varpi - z$ plane. This gives the SPH equation

$$\frac{dh_i}{dt} = \frac{1}{2}\frac{h_i}{\rho_i}\sum_j m_j\left[(v_{\varpi,j} - v_{\varpi,i})\frac{\partial W_{ij}}{\partial \varpi_i} + (v_{z,j} - v_{z,i})\frac{\partial W_{ij}}{\partial z_i}\right] \tag{5.47}$$

Note that the 2-dimensional rather than the 3-dimensional divergence is used here as the smoothing lengths only depend on the number density of particles in the $\varpi - z$ plane. The smoothing lengths are initially set according to the mass density. This results in smaller number of neighbours near the edges, typically dropping by a factor 2. In addition, the method results in numerical errors. To compensate for this the smoothing lengths for individual particles are adjusted when the number of particles under the kernel gets outside certain tolerances

$$h_i^n = \frac{h_i^{n-1}}{2}\left[1 + \left(\frac{N_s}{N_i^{n-1}}\right)^{1/2}\right] \tag{5.48}$$

with N_s the desired number of particles under the kernel and N_i^{n-1} the actual number of particles under the kernel for particle i. Note that here the total number of particles under the kernel is used, not the weighted number of particles under the kernel as in 5.44. The average number of particles under the kernel for all particles is also tracked and the smoothing lengths for all particles are adjusted

if this deviates too much from N_s. Unfortunately, these adjustments can lead to precisely the kind of instabilities that were to be prevented by the method. For the number of particles that were used for the simulations presented in this thesis the method performs adequately but for simulations with large numbers of particles, which have larger density contrasts, unphysical shocking occurs.

5.9 Time stepping

Both models use a second order Runge-Kutta method. Positions, velocities and energies are updated according to

$$\mathbf{r}_i^{n+1} = \mathbf{r}_i^n + f\mathbf{v}_i^{n+1/2}\Delta t + (1-f)\mathbf{v}_i^n\Delta t \tag{5.49}$$

$$\mathbf{v}_i^{n+1} = \mathbf{v}_i^n + f\mathbf{a}_i^{n+1/2}\Delta t + (1-f)\mathbf{a}_i^n\Delta t \tag{5.50}$$

$$u_a^{n+1} = u_i^n + f\dot{u}_i^{n+1/2}\Delta t + (1-f)\dot{u}_i^n\Delta t \tag{5.51}$$

with $\mathbf{a}$ the total acceleration and $f = 0.996$ as in [95]. Positions and velocities at half time are found as

$$\mathbf{r}_i^{n+1/2} = \mathbf{r}_i^n + 0.5\mathbf{v}_i^n\Delta t + \frac{1}{8}\mathbf{a}_i^n\Delta t^2 \tag{5.52}$$

$$\mathbf{v}_i^{n+1/2} = \mathbf{v}_i^n + 0.5\mathbf{a}_i^n\Delta t \tag{5.53}$$

The time-step in the 3D code is chosen as in Thacker et al. [127]

$$\delta t = C \min_i(h_i / \max_j(|\mathbf{v}_i - \mathbf{v}_j|) \tag{5.54}$$

with C, the Courant factor chosen as 0.3. The time-step in the 2D code was chosen as in [84]

$$\delta t = C \min(\delta t_f, \delta t_s) \tag{5.55}$$

with a Courant factor C of 0.3 and with

$$\delta t_f = \min_i \left(\frac{h_i}{|\mathrm{f}_i|} \right)^{1/2} \tag{5.56}$$

where $|\mathrm{f}_i|$ the acceleration of particle i, and with

$$\delta t_s = \min_i \frac{h}{c_i + 0.6(\alpha c_i + \beta \max_j \mu_{ij})} \tag{5.57}$$

where c_i is the sound speed of particle i, α and β the artificial viscosity parameters in the standard artificial viscosity, and μ_{ij} the artificial viscosity parameter given by equation 5.27.

5.10 Units

The units in both models are the same as those used by Navarro, Frenk and White [93]. Specifically they are:

$$G = 1 \tag{5.58}$$

$$[\text{mass}] = 10^{10} m_u \quad \mathrm{M}_\odot \tag{5.59}$$

$$[\text{distance}] = 1 d_u \quad \text{kpc} \tag{5.60}$$

$$[\text{time}] = 4.71 \times 10^6 \left(\frac{d_u^3}{m_u}\right)^{1/2} \text{yr} \tag{5.61}$$

$$[\text{velocity}] = 207.4 \left(\frac{m_u}{d_u}\right)^{1/2} \quad \text{km s}^{-1} \tag{5.62}$$

$$[\text{density}] = 6.77 \times 10^{-22} \left(\frac{m_u}{d_u^3}\right) \quad \text{g cm}^{-3} \tag{5.63}$$

$$[\text{power}] = 5.75 \times 10^{43} \left(\frac{m_u}{d_u}\right)^{5/2} \quad \text{erg s}^{-1} \tag{5.64}$$

where m_u and d_u are input parameters of the code. By changing the values of m_u and d_u the size of the system can be scaled. For these simulations $m_u = 100$ and $d_u = 100$, which gives a system of $10^{12}\text{M}_\odot$.

5.11 Initial Conditions

A galaxy of $10^{12}\text{M}_\odot$ is simulated, starting at its turn around radius of 100 kpc. Initially the mass is homogeneously distributed, with the particles located at equal distances from each other and the masses of the rings in the 2D simulations proportional to ϖ. In the 3D simulations 2416 particles are used, in the 2D simulations 1000 particles. The number of gas and dark matter particles is equal in all simulations. The gas fraction is can be varied but is taken as 0.2 for most simulations. The sphere is given a rigid rotation, such that the spin parameter $\lambda = 0.08$ [46].

Initial conditions for the formation of a single galaxy are not well defined. The ideal way of creating the initial conditions would be by doing a full cosmological simulation in which the proto-galaxy is formed by mergers of smaller systems and affected by tidal forces of surrounding systems [91]. However, the resolution of these kind of simulations is limited and single galaxies are often represented by a very small number of particles. An alternative approach is to take initial conditions in a similar way to cosmological simulations, imposing small scale fluctuations according to a prescribed power spectrum [125]. Here velocity perturbations of the form

$$\Delta v_{\varpi,i} = \Sigma_{j=1}^{10} v_{pt} \left[\sin\left\{\left(j\varpi_i + 2a(j)\right)\pi\right\} + \sin\left\{\left(jz_i + 2b(j)\right)\pi\right\}\right] \tag{5.65}$$

were imposed on the initially uniform stationary particles. The parameters $a(j)_i$ and $b(j)_i$ are random numbers between 0 and 1, and v_{pt} is an input parameter. Velocity perturbations in other directions were calculated in a similar way.

Chapter 6

The In-shock Cooling Problem

6.1 Introduction

The in-shock cooling problem was first mentioned by Nulsen and Fabian [99] and a first attempt to find a solution was made by Maguire [78]. The essence of the problem is simple: gas spends longer in simulated shocks than in real shocks and therefore has more time to cool. This can prevent it from ever reaching the correct post-shock temperature. The problem has largely been ignored in the literature [100], even though possibly related problems, such as small disc sizes and dense knots of gas in the center of the galaxy are well known [124] [41] [91]. Hutchings and Thomas [65] did shock tube simulations with the shock crossing time of the same order as the cooling time, and found that the gas did not reach the correct post-shock temperature.

An estimate of the ratio of the shock thickness in simulations to the shock thickness in reality was given by Maguire [78]. Real shocks are a few particle mean free path lengths thick [115]. For a gas at $10^6 T_6$ K with electron density $n_e\,\mathrm{cm}^{-3}$, this is roughly $10^{-3}\,T_6^2/n_e$ pc ([121]). Post-shock temperatures and densities are of order 10^6 K and $0.1\,\mathrm{cm}^{-3}$, respectively, giving a shock thickness of about 10^{-2} pc. In SPH simulations, shocks are smoothed by artificial viscosity over several smoothing lengths. In a typical galaxy collapse simulation, smoothing lengths in well resolved areas are of the order of 100 pc to 1 kpc. Thus, the ratio of the simulated to the actual shock thickness is of the order of 10^4 to 10^5 or more. Since the speed of shock propagation is not severely affected by spatial resolution, the time necessary to pass through the shock is increased by a similar factor. This means that, while cooling during the shock is negligible in reality, the gas can cool considerably in simulated shocks.

6.2 Description of the Problem

The in-shock cooling problem will now be treated in more detail by considering a steady 1-dimensional flow. In this case the continuity equation becomes

$$\frac{d}{dx}(\rho v) = 0, \tag{6.1}$$

where ρ is the density and v the velocity. The momentum equation is

$$\frac{d}{dx}(\rho v^2 + P - T_{xx}) = 0, \tag{6.2}$$

where P is the pressure and T_{xx} the x-x component of the viscous stress tensor, and the energy equation may be written as

$$\frac{d}{dx}\left[\rho v(H + \frac{1}{2}v^2) - T_{xx}v\right] = -\rho^2\Lambda, \tag{6.3}$$

where $H = \gamma P/[(\gamma - 1)\rho]$ is the specific enthalpy and $\rho^2\Lambda$ the power radiated per unit volume.

For a real radiative shock there are four regions of interest ([115], for a review of interstellar shocks see [42]). In the upstream region the gas is cold and moves at supersonic speeds relative to the shock front. The gas is then suddenly decelerated, compressed and heated, which happens in a thin viscous transition layer where $T_{xx}v$ is large. The time the gas takes to cross this region is short compared to its cooling time and cooling can be ignored, resulting in the energy equation

$$\frac{d}{dx}\left[\rho v(H + \frac{1}{2}v^2) - T_{xx}v\right] = 0 \tag{6.4}$$

The conditions immediately downstream from this region are given by the Rankine-Hugoniot jump conditions

$$\frac{\rho}{\rho_0} = \frac{v_0}{v} = \frac{(\gamma+1)M_0^2}{(\gamma+1)+(\gamma-1)(M_0^2-1)} \tag{6.5}$$

$$\frac{P}{P_0} = \frac{(\gamma+1)+2\gamma(M_0^2-1)}{(\gamma+1)} \tag{6.6}$$

$$\frac{T}{T_0} = \frac{[(\gamma+1)+2\gamma(M_0^2-1)][(\gamma+1)+(\gamma-1)(M_0^2-1)]}{(\gamma+1)^2 M_0^2} \tag{6.7}$$

with ρ_0, P_0 and T_0, the density, pressure and temperature in the undisturbed upstream region respectively, γ the ratio of the specific heats and $M_0 = v_0/\sqrt{\gamma T_0}$ the Mach number of the upstream gas. The result is hot gas, moving subsonically relative to the shock front. The gas then enters the radiative relaxation region, where it cools by emitting radiation. This causes the gas to compress and further decelerate. T_{xx} is negligible in this region, and it can be shown that

$$\frac{1}{(\gamma-1)}(c^2 - v^2)\frac{dv}{dx} = -\rho\Lambda \tag{6.8}$$

with c the sound speed of the gas $c^2 \equiv \gamma P/\rho$. Finally, in the downstream region, the gas is cold again and moves very slowly.

For a numerical shock, when in-shock cooling is present, the situation is more complicated. The artificial viscosity spreads the viscous transition layer over a much larger distance than in a real shock and the time to cross this region is no longer short in comparison to the cooling time of the gas. There is no clear distinction any more between the viscous transition layer and the radiative

relaxation layer. The gas is radiating its energy away while it is still being heated, resulting in a much lower post-shock temperature than given by equation 6.7. Both T_{xx} and $\rho^2\Lambda$ are appreciable in the energy equation 6.3. Integrating equation 6.1, the continuity equation is

$$\rho v = \rho_0 v_0 \tag{6.9}$$

and the momentum equation can be integrated to give

$$\rho v^2 + P - T_{xx} = \rho_0 v_0^2 + P_0 \tag{6.10}$$

where it is assumed that T_{xx} is negligible in the upstream flow. This can be solved for T_{xx}. Eliminating T_{xx} from the energy equation rearranging terms and using equation 6.9 gives

$$\rho_0 v_0 \frac{dE}{dx} = -\rho^2\Lambda - [P_0 + \rho_0 v_0(v_0 - v)]\frac{dv}{dx} \tag{6.11}$$

Since the specific energy, $E = H - P/\rho$, is proportional to the temperature, the shock fails, i.e. the temperature of the gas does not rise any further, when the RHS is negative (assuming the flow is to the right). This equation shows the effect of poor spatial resolution. Lower resolution reduces values for dv/dx, as the shock is spread over a larger distance, while $\rho^2\Lambda$ is essentially unchanged and so becomes increasingly important as the shock becomes thicker. This causes a thicker shock to fail at an earlier stage. Thoul and Weinberg [131] [132] performed 1-dimensional simulations, including the effect of radiative cooling, and found that a smaller fraction of the gas cooled catastrophically in their simulation than

in 3-dimensional simulations. This may well have been caused by the higher resolution in their models.

dv/dx will also depend on the form of the artificial viscosity used. Thacker et al. [127] did simulations of the formation of a disc galaxy, similar to the ones described in this chapter, with 12 different combinations of artificial viscosity and symmetrisation. They found that minor differences in the form of the artificial viscosity or the symmetrisation used could double the amount of hot gas produced.

The form of the cooling function used also determines whether the shock will fail or not. Kay et al. [72] compared simulations of galaxy formation with a cooling function consisting of a series of power-law fits to a to cooling function based on tabulated values and found an increase in galaxy masses of 10%. While this effect is not necessarily due to differences in the in-shock cooling, it can be argued that, as with artificial viscosity, minor differences in the form of the cooling function can have a large influence on the outcome of simulations because they can determine whether shocks fail or not.

Finally, it should be mentioned that discreteness of the numerical simulations means that in some cases equation 6.11 will lead to incorrect results. Strong shocks and relatively long time-steps may cause the gas to 'jump over' the critical point where the shock would fail. Since Λ is relatively large at low temperatures the shock is most likely to fail when the gas just enters the shock. In some cases the analytical condition might predict that the shock would fail while this does not happen in the simulation. The implication of this is that, contrary to what would intuitively be expected, taking shorter time-steps can make the in-shock

cooling problem worse. It is expected, however, that in most cases the analytical description gives a good indication of what happens in the simulations.

6.3 Demonstration of the Problem

To investigate the effect of in-shock cooling in simulations of galaxy formation, the viscous heating rate, the radiative cooling rate and the change in thermal energy per unit time due to the pressure forces were tracked separately in both the 2D and the 3D collapse models. This is done by splitting the energy equation into three separate parts. The viscous heating rate per unit mass is calculated as

$$\begin{aligned} \varepsilon_{\mathrm{vis,i}} &= \frac{1}{2}\sum_j m_j \Pi_{ij} \\ &\quad \times \left[\left(\frac{\varpi_j}{\varpi_i} v_{\varpi,j} - v_{\varpi,i} \right) \frac{\partial W_{ij}}{\partial \varpi_i} + (v_{z,j} - v_{z,i}) \frac{\partial W_{ij}}{\partial z_i} \right] \end{aligned} \tag{6.12}$$

for the 2D model and

$$\varepsilon_{\mathrm{vis,i}} = \frac{1}{2}\sum_j m_j \Pi_{ij} \mathbf{v}_{ij} \cdot \nabla_i W_{ij} \tag{6.13}$$

for the 3D model, where Π_{ij} is given by equation 5.26 for both models, $\mathbf{v}_{ij} = \mathbf{v}_i - \mathbf{v}_j$ and W_{ij} the kernel. The change in thermal energy due to the pressure force, i.e. the adiabatic compression, is

$$\begin{aligned} \varepsilon_{\mathrm{pres,i}} &= \frac{1}{2}\sum_j m_j \left(\frac{P_i}{\rho_i^2} + \frac{P_j}{\rho_j^2} \right) \\ &\quad \times \left[\left(\frac{\varpi_j}{\varpi_i} v_{\varpi,j} - v_{\varpi,i} \right) \frac{\partial W_{ij}}{\partial \varpi_i} + (v_{z,j} - v_{z,i}) \frac{\partial W_{ij}}{\partial z_i} \right] \end{aligned} \tag{6.14}$$

for the 2D model and

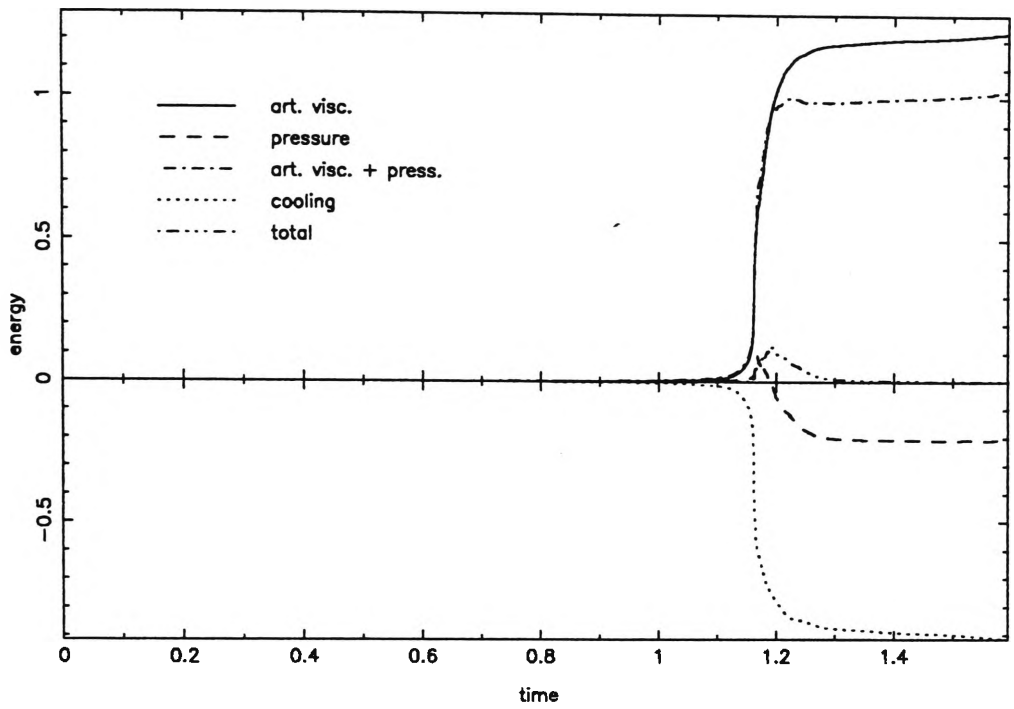

Figure 6.1: Heating and cooling in the 2D model

$$\varepsilon_{\text{pres,i}} = \frac{1}{2}\sum_j m_j \left(\frac{P_i}{\rho_i^2} + \frac{P_j}{\rho_j^2}\right) \mathrm{v}_{ij} \cdot \nabla_i W_{ij} \tag{6.15}$$

for the 3D model, where P_i is the pressure for particle i and j. Radiative cooling is found by

$$\varepsilon_{\text{rad,i}} = \rho_i k \Lambda \tag{6.16}$$

with Λ the cooling function and k a conversion factor, to convert the cooling rate per unit mass to program units.

$\varepsilon_{\text{vis,i}}$, $\varepsilon_{\text{pres,i}}$ and $\varepsilon_{\text{rad,i}}$ were integrated using the Runge-Kutta method described in section 5.9 to obtain the total viscous heating, total adiabatic change in thermal energy due to pressure and total energy radiated away. These are plotted against time in Figure 6.1 for the 2D model with gas fraction 0.2. As can be seen, most of the thermal energy gained by viscous heating is radiated away almost

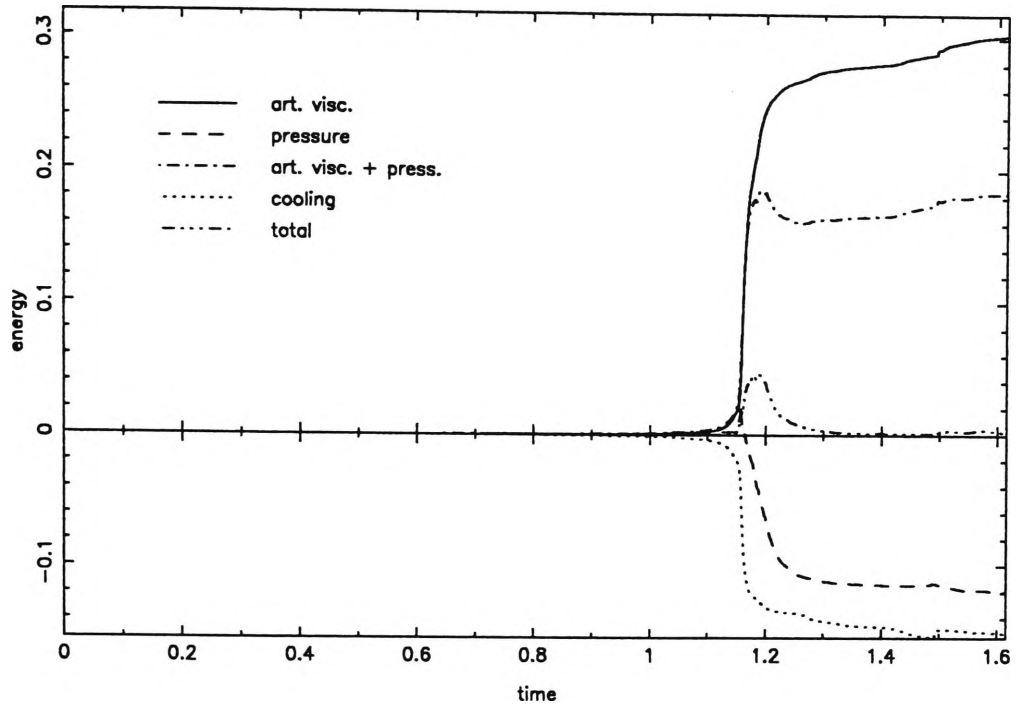

Figure 6.2: Heating and cooling in the 2D model for a gas fraction of 0.05

simultaneously and little thermal energy is left. If a gas fraction of 0.05 is used instead (see Figure 6.2), the effect is a bit smaller, but still very clear. The gas reaches a slightly higher temperature, but most of the energy from viscous heating is still radiated away almost immediately.

The difference between Figure 6.1 and Figure 6.2 can be explained by the density. The lower gas fraction results in lower density for the gas. Radiative cooling per unit mass is proportional to the density, so thermal energy is radiated away less quickly if the gas fraction is lower.

For the 3D model the effect of cooling during the shock is even stronger than for the 2D model. Both the viscous heating rate and the radiative cooling rate are lower than in the 2D model, but they continue much longer. Energy is radiated away so efficiently that the amount of hot gas produced is practically negligible. The difference between the 2D and 3D model is mainly caused by the difference

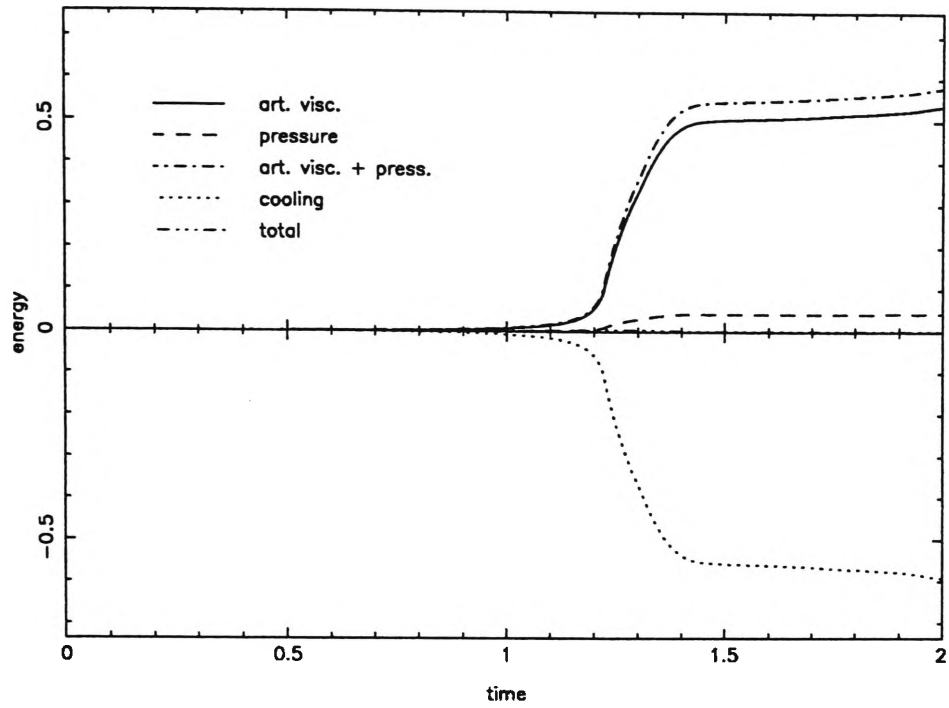

Figure 6.3: Heating and cooling in the 3D model

in spatial resolution. The 2D model contains 1000 particles, corresponding to $\sim$ 30000 particles in a 3 dimensional simulation for the same spatial resolution. The 3D model has only 2416 particles, however, so the resolution is much lower in the 3D than in the 2D model. Shocks are therefore more extended in the 3D model than in the 2D model.

Although Figure 6.1, 6.2 and 6.3 clearly show that viscous heating and radiative cooling occur simultaneously, they do not prove the existence of in-shock cooling. If some particles, which have already passed through the shock, are cooling while other particles, which are still crossing the shock, are heating, the rates may balance without in-shock cooling. In-shock cooling occurs when particles cool while passing through the shock, which means that they cool while they are still heating. In order to investigate whether the effect seen in Figure 6.1 - 6.3 is genuinely in-shock cooling $\varepsilon_{\rm pres}$ and $\varepsilon_{\rm rad}$ were saved at regular intervals for all

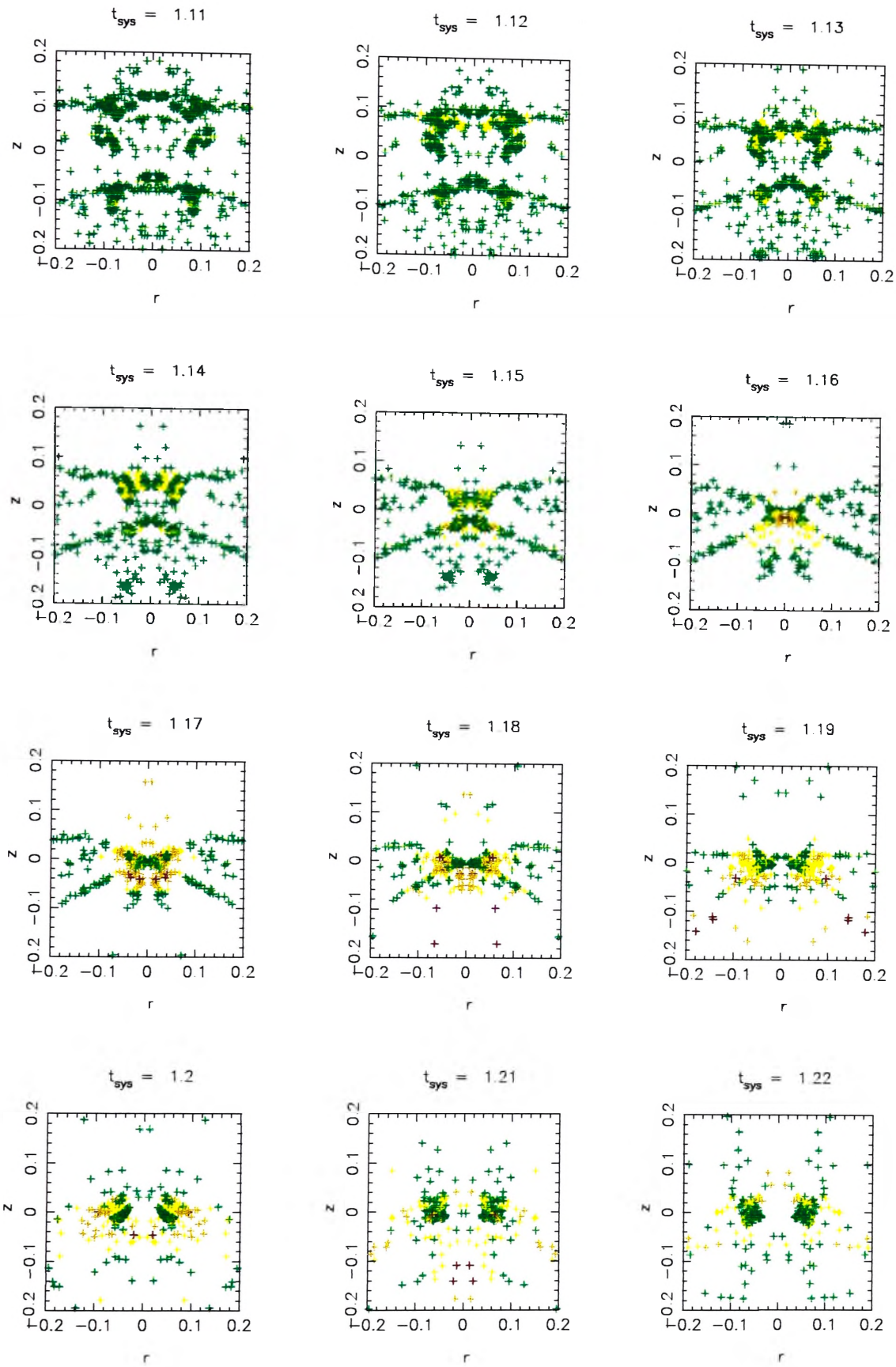

Figure 6.4: Viscous heating in the 2D model

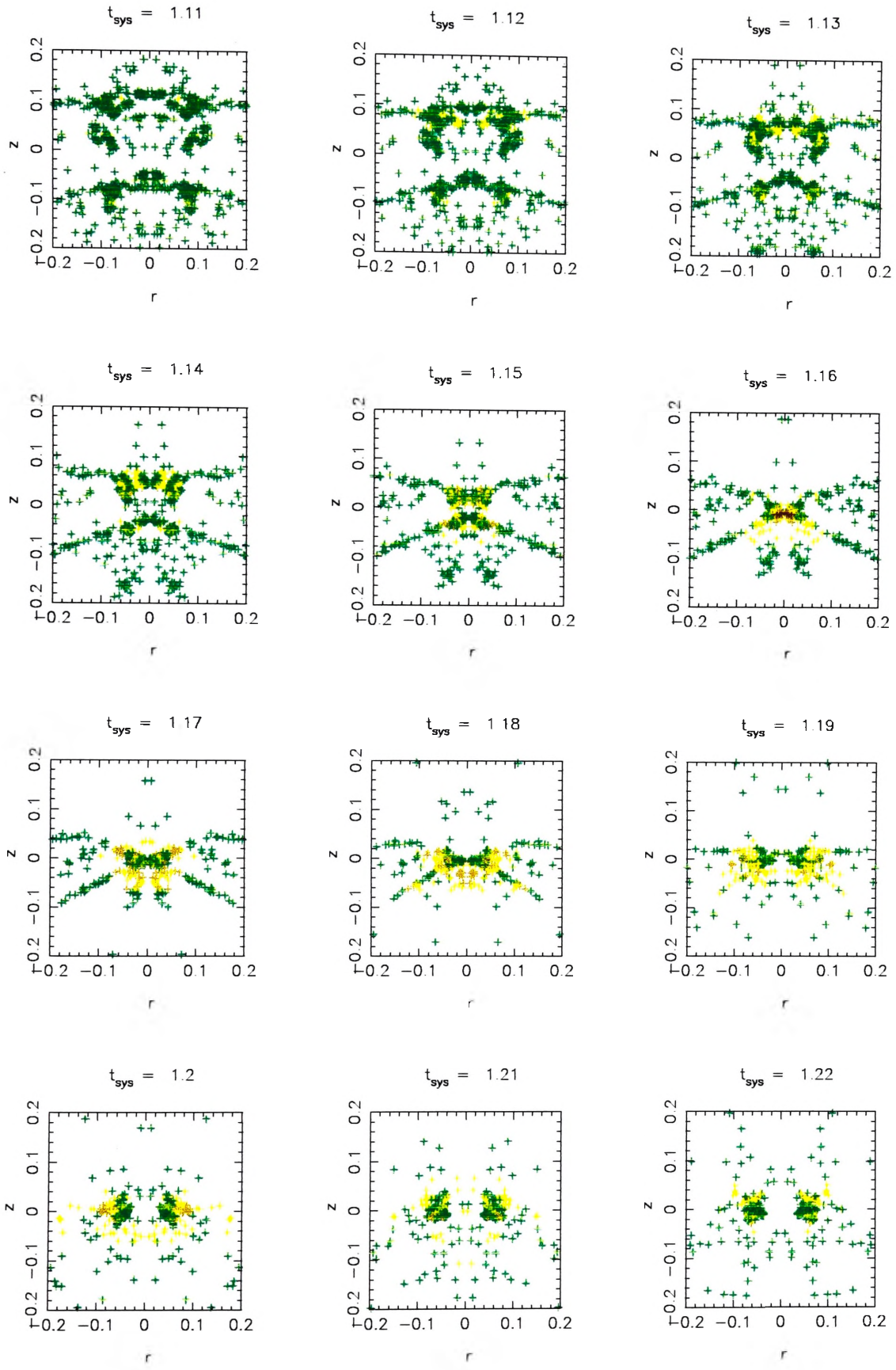

Figure 6.5: Radiative cooling in the 2D model

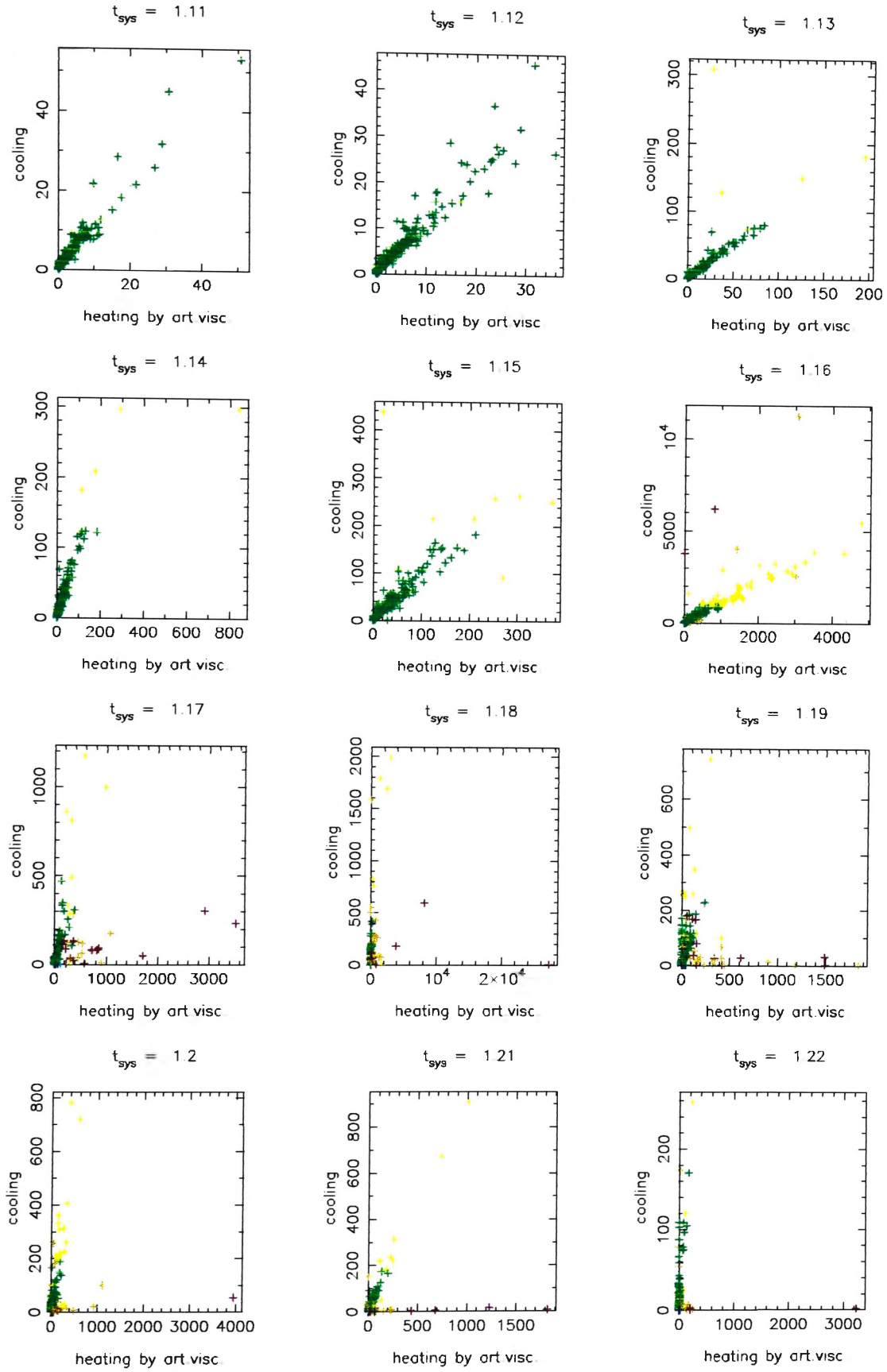

Figure 6.6: Radiative cooling vs viscous heating in the 2D model

particles. Figures 6.4 and 6.5 show the positions of the particles at different times during the 2D simulation with the colours of the particles indicating the amount of viscous heating and radiative cooling, respectively. The same colors represent the same power per unit mass in both plots, with units as given in section 5.10

green: $\varepsilon < 10$

yellow: $10 < \varepsilon < 10^2$

orange: $10^2 < \varepsilon < 10^3$

red: $10^3 < \varepsilon < 10^4$

purple: $\varepsilon > 10^4$

Notice the strong similarity of Figures 6.4 and 6.5.

Figure 6.6 shows plots of radiative cooling vs. heating rates per unit mass for the same times during the simulation. It can be seen that viscous heating and radiative cooling are correlated, indicating in-shock cooling.

Figure 6.7 and Figure 6.8 show the viscous heating and radiative cooling in the 3D model. The plots give the top view of the proto-galaxy at different stages during the collapse, with colours again indicating viscous heating and radiative cooling rates, using the same scale as in Figure 6.4. Both the viscous heating rate and the radiative cooling rate are on average lower than in the 2D model. The similarities in colour between Figure 6.7 and Figure 6.8 indicate that viscous heating and radiative cooling are of the same order of magnitude for most particles. Plots of radiative cooling vs viscous heating (Figure 6.9) show a stronger the correlation between viscous heating and radiative cooling than for the 2D model.

The strong correlation between viscous heating and radiative cooling in the

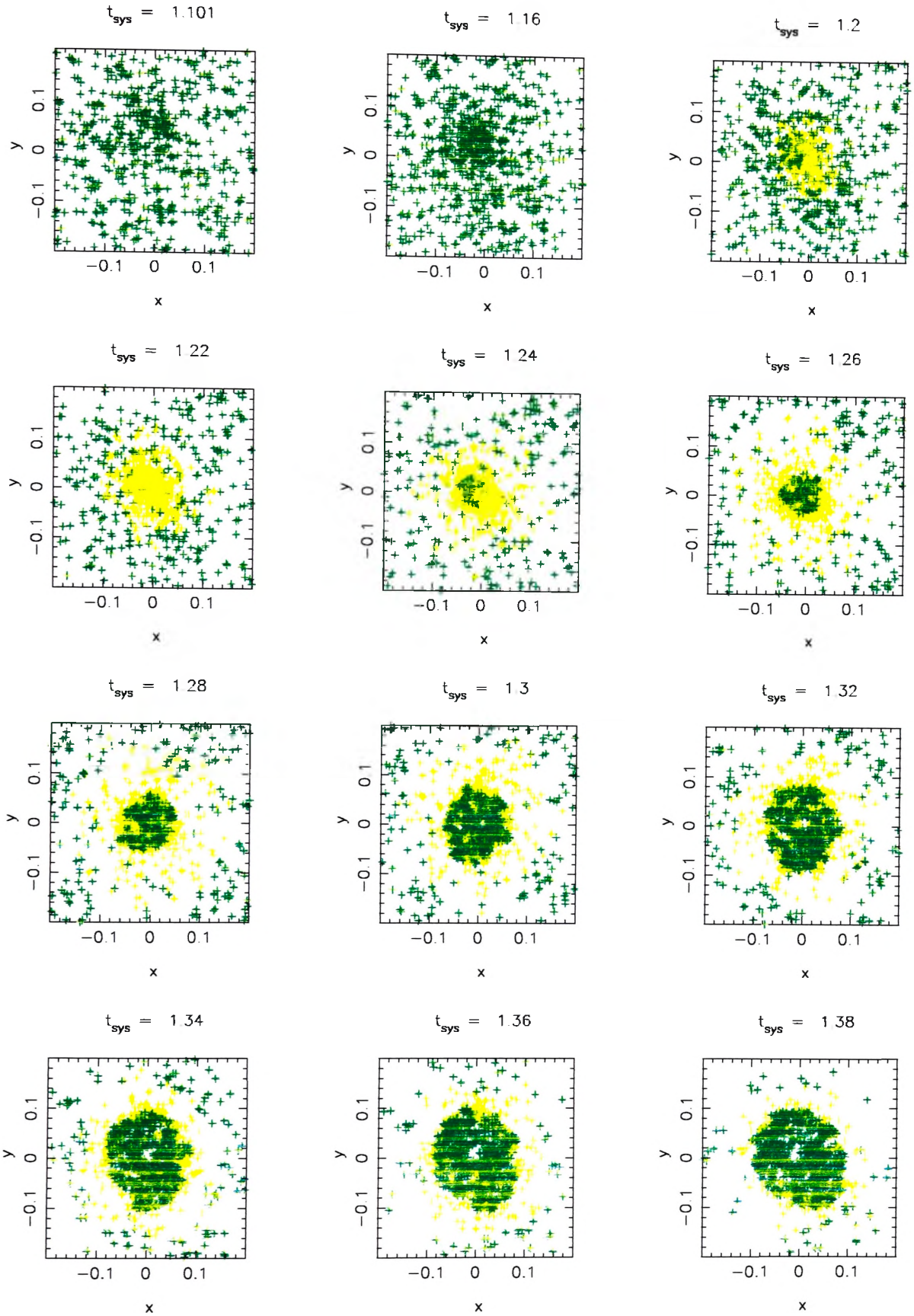

Figure 6.7: Viscous heating in the 3D model

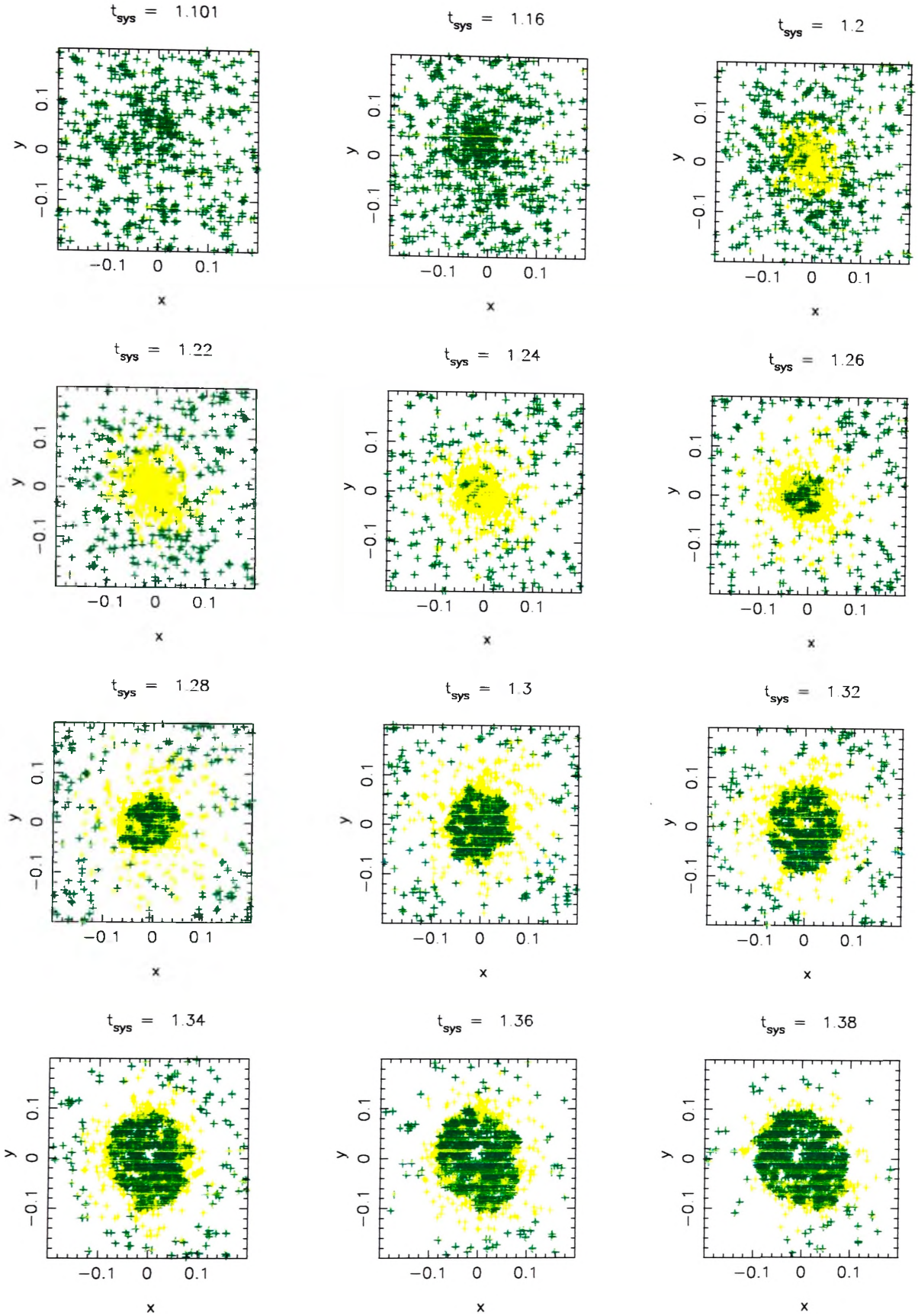

Figure 6.8: Radiative cooling in the 3D model

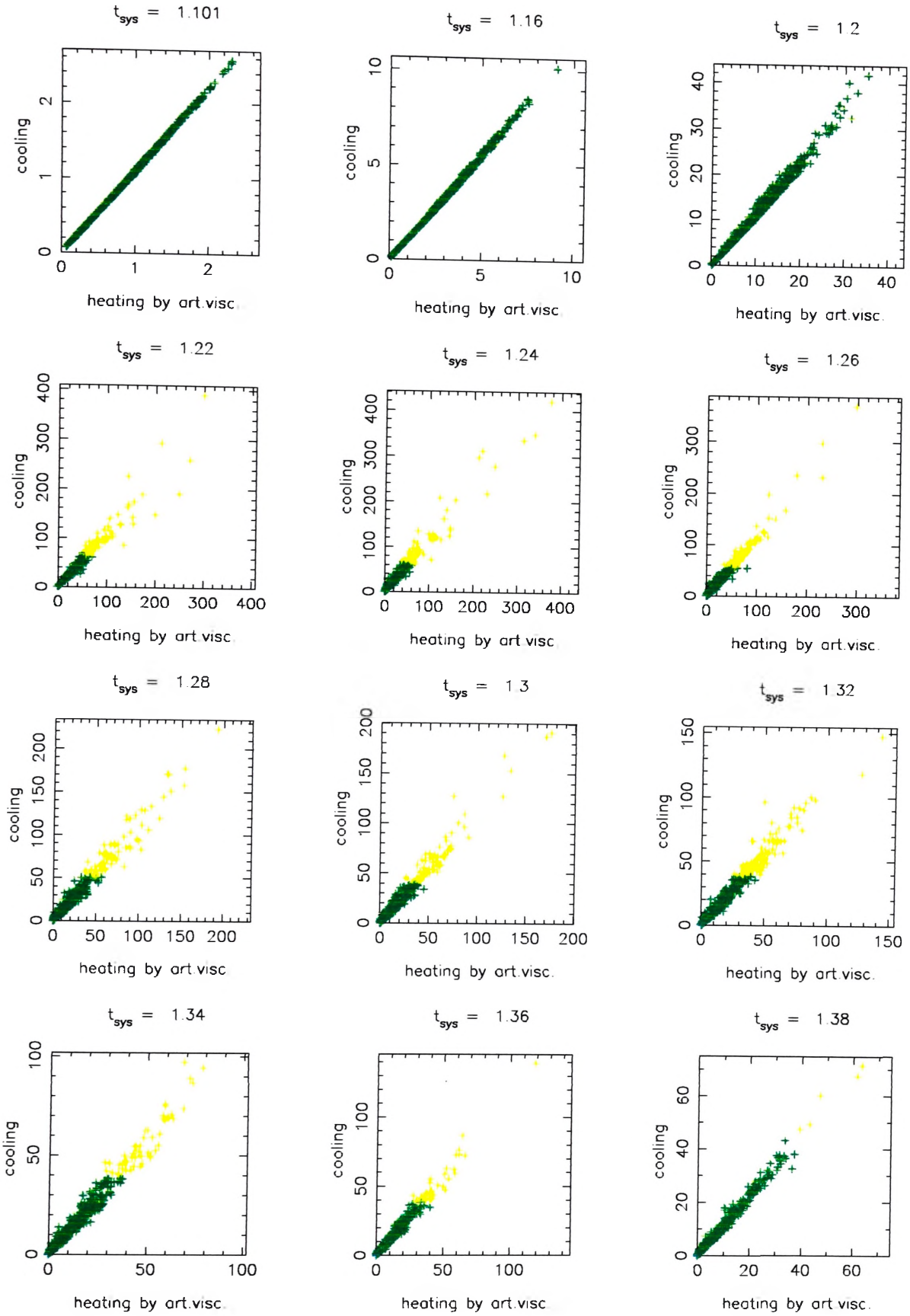

Figure 6.9: Radiative cooling vs viscous heating in the 3D model

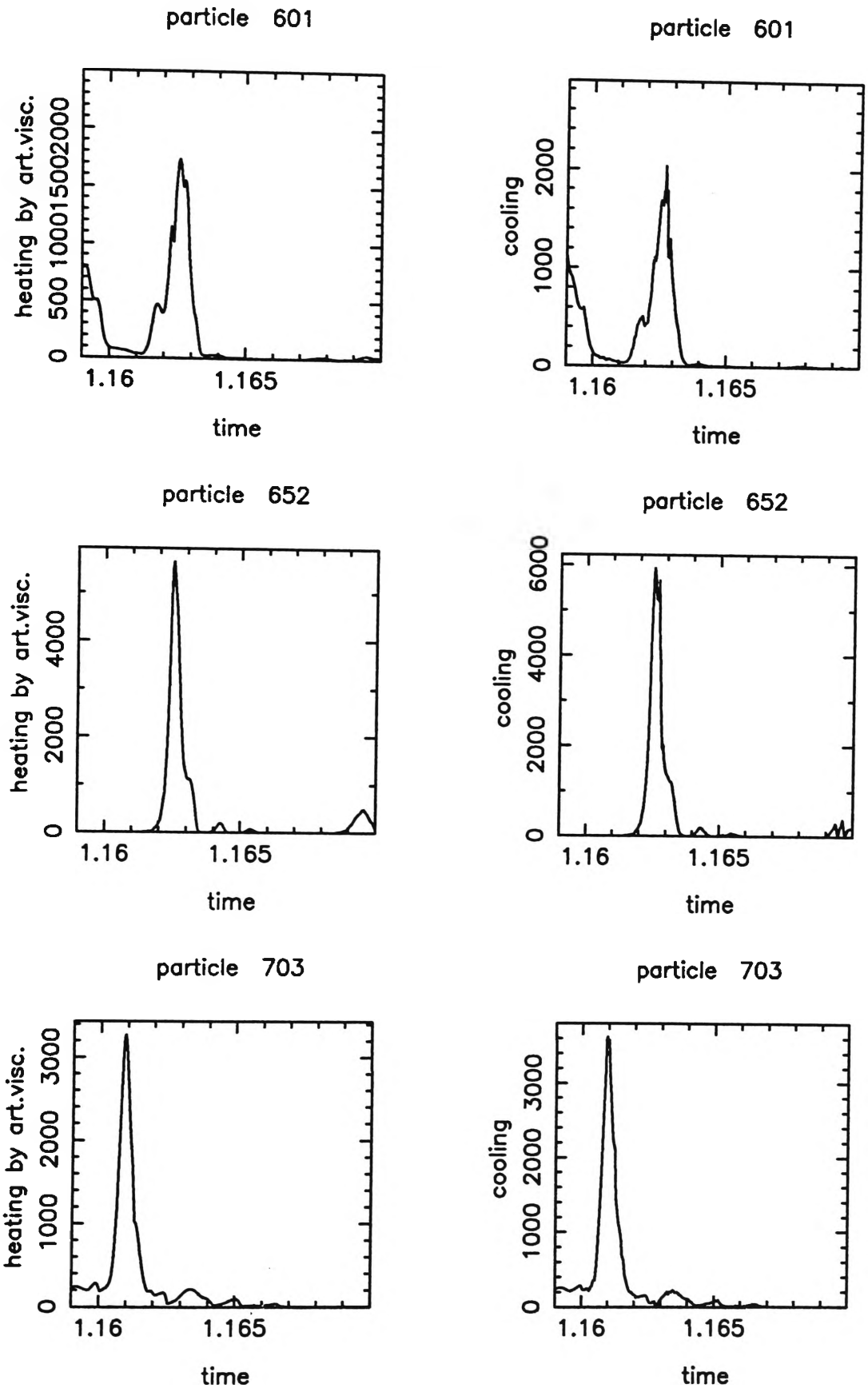

Figure 6.10: Viscous heating and radiative heating vs time for individual particles

2D model is illustrated further in Figure 6.10. For a few particles viscous heating and radiative cooling were tracked at every step and plotted against the time. The results show that for most of these particles, cooling tracks viscous heating with remarkable accuracy. The thermal energy gained by viscous heating is mostly radiated away in the same time-step.

6.4 Possible Solutions for the In-shock Cooling Problem

6.4.1 Requirements

A number of considerations need to be kept in mind when searching for a solution for the in-shock cooling problem. First and most important, the solution should actually reduce cooling in shocks. This means that it results in significantly less cooling for particles that are undergoing shock heating, resulting in higher post-shock temperatures. Second, the solution should not interfere with other aspects of the code, resulting in unphysical behaviour of the model. In particular, cooling should not be reduced significantly in particles which are not shocking. Third, the solution should not be costly in terms of computation time or storage capacity. It should therefore not require a large amount of extra calculations or data. It is also desirable that a good solution is be easy to implement. Fourth, in order to be widely applicable, the solution should only involve dimensionless criteria. Fifth, the solution should be universal, that is it should not depend on what is simulated, nor on details of the implementation of a flow model. The solution should be valid for all types of shocks, independent of the Mach number or the pre-shock temperature, density and velocity. The solution should

also be independent of the form of the artificial viscosity or the cooling function and should be unaffected by the resolution of the code or the number of spatial dimensions. Finally, the solution should be independent of reference frame.

6.4.2 Proposed Methods

The in-shock cooling problem is caused by the fact that shocks in simulations are much thicker than in reality and the gas therefore has more time to cool while traveling through them. Although increasing the resolution reduces the shock thickness, it is not feasible to increase the resolution to the extent that in-shock cooling becomes irrelevant. Possible solutions for the problem are therefore focused on reducing the amount of cooling. Maguire [78] mentions a few possible solutions, all of which he found unsatisfactory. These solutions will not be investigated further here, but they are mentioned for the sake of completeness.

1. Increasing the temperature used for the cooling function. Because cooling is very efficient at low temperatures, using a higher temperature in the cooling function, $T + q$ instead of T, can reduce cooling during the shock. If q could be adjusted in such a way that it is 0 before the shock and that $T + q$ equals the correct post-shock temperature during the shock, this might be a valuable approach. However, this amounts to making the simulated shocks infinitesimally thick and no way has been found to achieve this yet. Attempts to implement this method are likely lead to incorrect cooling rates before and after the shock. Maguire used the increase in thermal energy during the current time-step due to viscous heating and adiabatic compression for q and found that this actually made the in-shock cooling problem worse. The cooling rate increases rapidly with

temperature for temperatures just above 10^4 K, and using $\Lambda(T+q)$, therefore leads to more cooling as particles enter the shock.

2. Adjusting the initial conditions to reduce the amount of cooling. Since the cooling time is inversely proportional to the density, cooling can be reduced by lowering the density of the gas. This can be done by reducing the gas fraction or by increasing the initial radius of the gas sphere $R_{\rm gas}$. As could be seen from Figure 6.2, a reduced gas fraction does result in more hot gas, but the in-shock cooling problem is still quite severe. Maguire found that increasing $R_{\rm gas}$ also did not prevent in-shock cooling. There are two fundamental problems with this approach. First, reducing the density of the gas lowers the amount of cooling overall, not specifically the amount of cooling in shocks. It therefore does not fulfil the requirement, mentioned above, that the solution should not interfere with other aspects of the model. Second, the choice of initial conditions should be motivated by what is physically most realistic. A good solution for the in-shock cooling problem should not require adjustment of the physical properties of the simulated system.

3. Turning cooling off during the initial part of the simulation. Cooling can be turned off until shocks have propagated through a significant part of the collapse simulation. Again, this solution fails the requirement that it should not interfere with other aspects of the model, since cooling is also turned off in situations where it would be physically realistic. The time when cooling is turned on again has to be chosen carefully. Turning cooling on too early will lead to in-shock cooling. Turning it on too late prevents some physically realistic cooling after the shock.

A good solution for the in-shock cooling problem will reduce the amount

of cooling for particles in shocks while not reducing it for other particles. Two methods which attempt to achieve this are investigated here: reducing the amount of radiative cooling by the amount of viscous heating and turning cooling off in shocks.

For the first method, the viscous heating rate is calculated according to 6.12 and 6.13, for the 2D and 3D model respectively, and the uncorrected radiative cooling rate is calculated according to 6.16. The cooling rate is then reduced by the viscous heating rate, i.e. the cooling rate per unit mass used in the model is taken as

$$\varepsilon_{\mathrm{rad}*,\mathrm{i}} = \varepsilon_{\mathrm{rad,i}} - \varepsilon_{\mathrm{vis,i}} \tag{6.17}$$

Since the viscous heating rate and the radiative cooling rate are of the same magnitude in shocks, based on the strong correlation between them (see Figures 6.6 and 6.9), it is expected that this method will reduce the cooling in shocks significantly. The method does not interfere with physically realistic cooling outside shocks, because cooling will not be reduced significantly if the viscous heating rate is small. The method does require splitting the energy equation into separate terms, but this is not difficult to implement and it does not result in a large number of extra calculations. Since the dimensions of the viscous heating rate and the radiative cooling rate are the same, the method can be applied irrespective of the dimensions used in the code. If the cooling rate is larger than the viscous heating rate, some in-shock cooling still occurs so the method may not resolve the in-shock cooling problem completely. In those cases where the viscous heating rate is larger than the cooling rate, cooling is set to zero.

The second method, switching off cooling in shocks, is the obvious solution to the in-shock cooling problem, but the method is not without its difficulties. The main difficulty of this method is that it is necessary to decide whether a particle is in a shock or not. The value of the method depends on whether or not a good criterion can be found to locate the shocks. If such a criterion can be found, the method seems promising.

6.5 Reduce Cooling by Amount of Viscous Heating

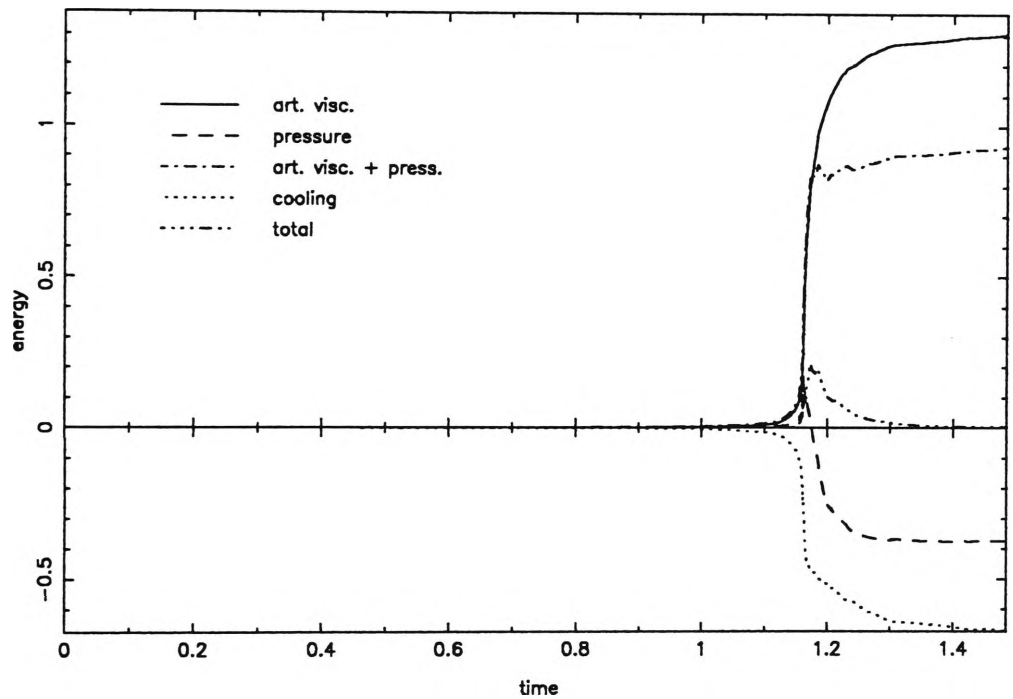

Figure 6.11: Heating and cooling in the 2D model with cooling reduced by the amount of viscous heating

In both the 2D and the 3D models the amount of radiative cooling was reduced by the amount of viscous heating. As shown in Figures 6.11 and 6.12, the modification slightly increases amount of hot gas in the 2D model, but has little effect on the 3D model. Figures 6.13 and 6.14 show plots of the positions of the

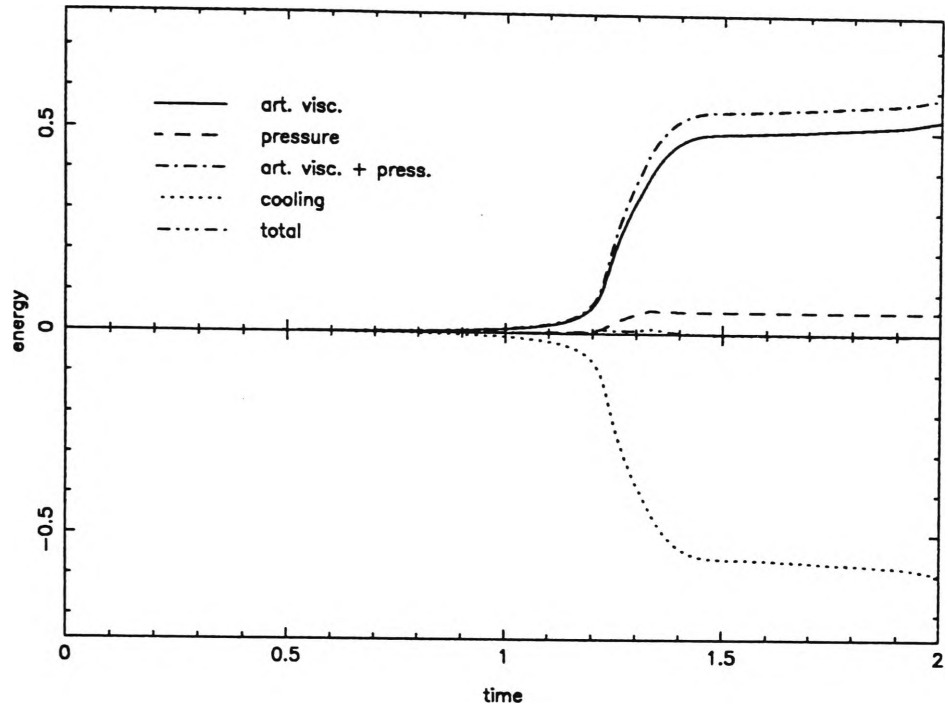

Figure 6.12: Heating and cooling in the 3D model with cooling reduced by the amount of viscous heating

particles at different times during the 3D simulation, with the colors this time indicating the temperature:

dark blue $T < 10^3$K

light blue: $10^3\text{K} < \text{T} < 10^4\text{K}$

green: $10^4\text{K} < \text{T} < 10^5\text{K}$

yellow: $10^5\text{K} < \text{T} < 10^6\text{K}$

orange: $10^6\text{K} < \text{T} < 10^7\text{K}$

red: $10^7\text{K} < \text{T} < 10^8\text{K}$

purple: $T > 10^8$K

A comparison of Figure 6.13 and Figure 6.14 shows that the temperature of the gas is slightly increased, when cooling is reduced by the amount of viscous heating, but this effect lasts only for a short time. In the 2D model, the modification has a stronger effect (Figures 6.15 and 6.16), but it is still insufficient to provide a

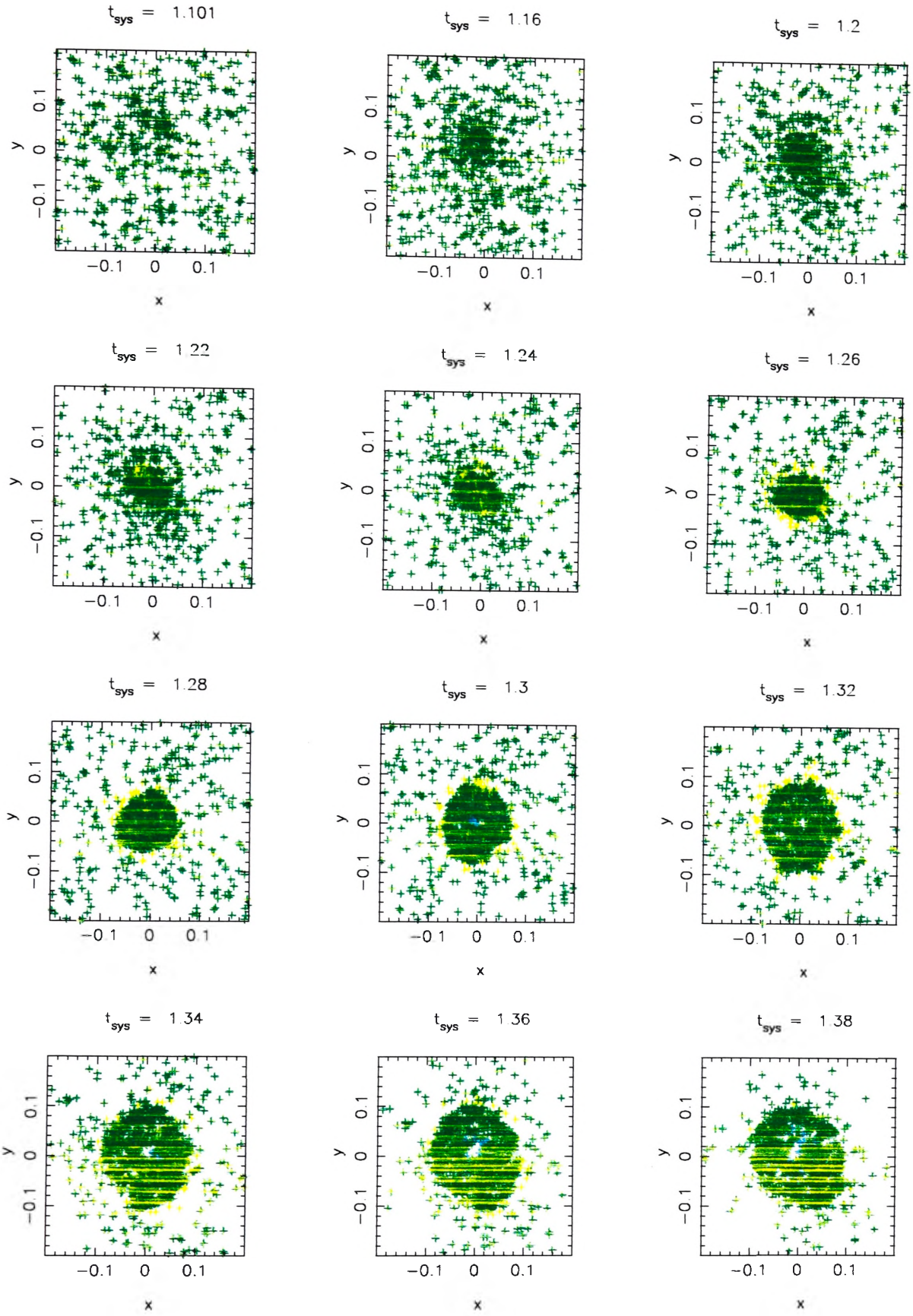

Figure 6.13: Temperature in the 3D model

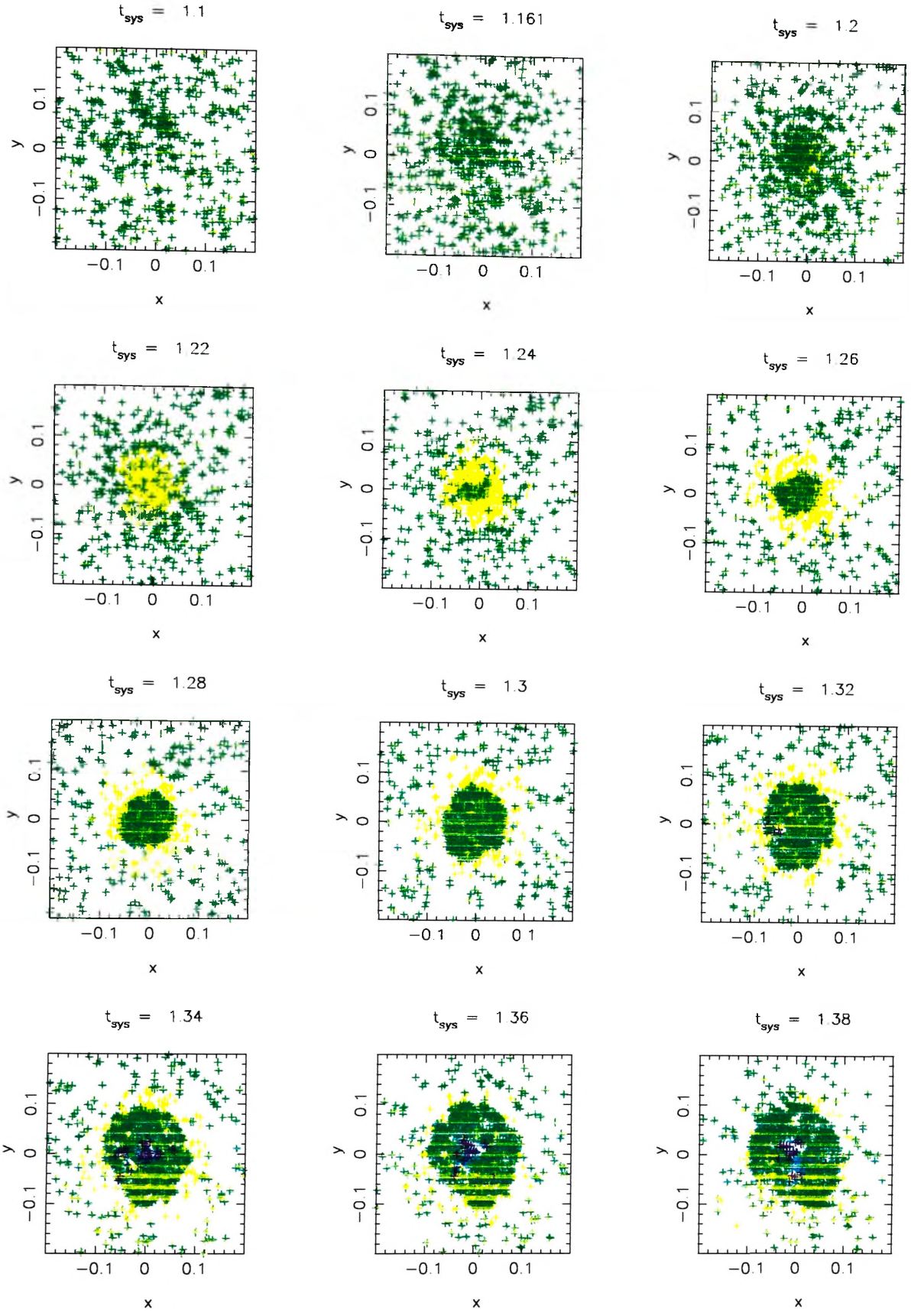

Figure 6.14: Temperature in the 3D model when cooling is reduced by the amount of viscous heating

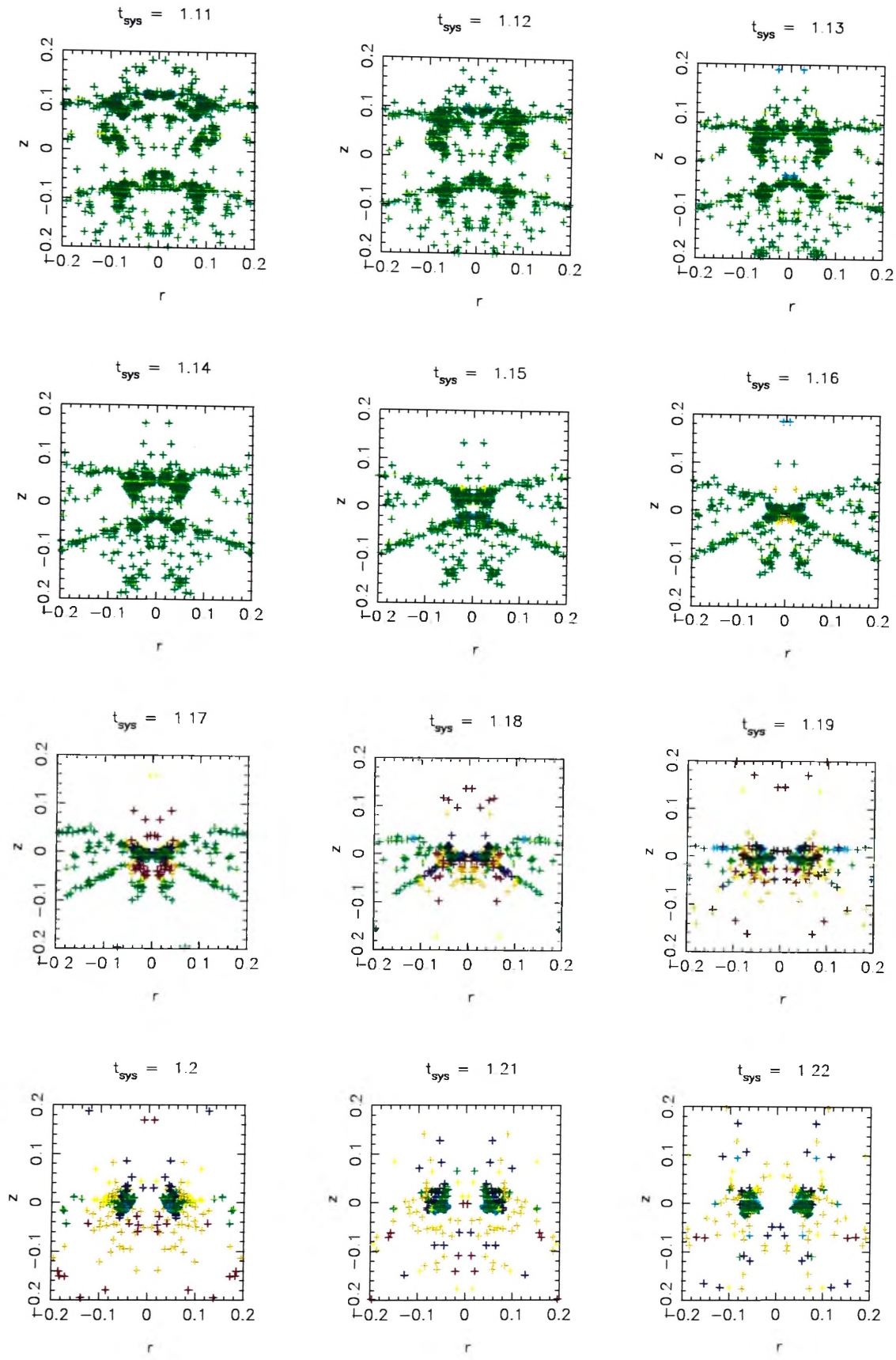

Figure 6.15: Temperature in the 2D model

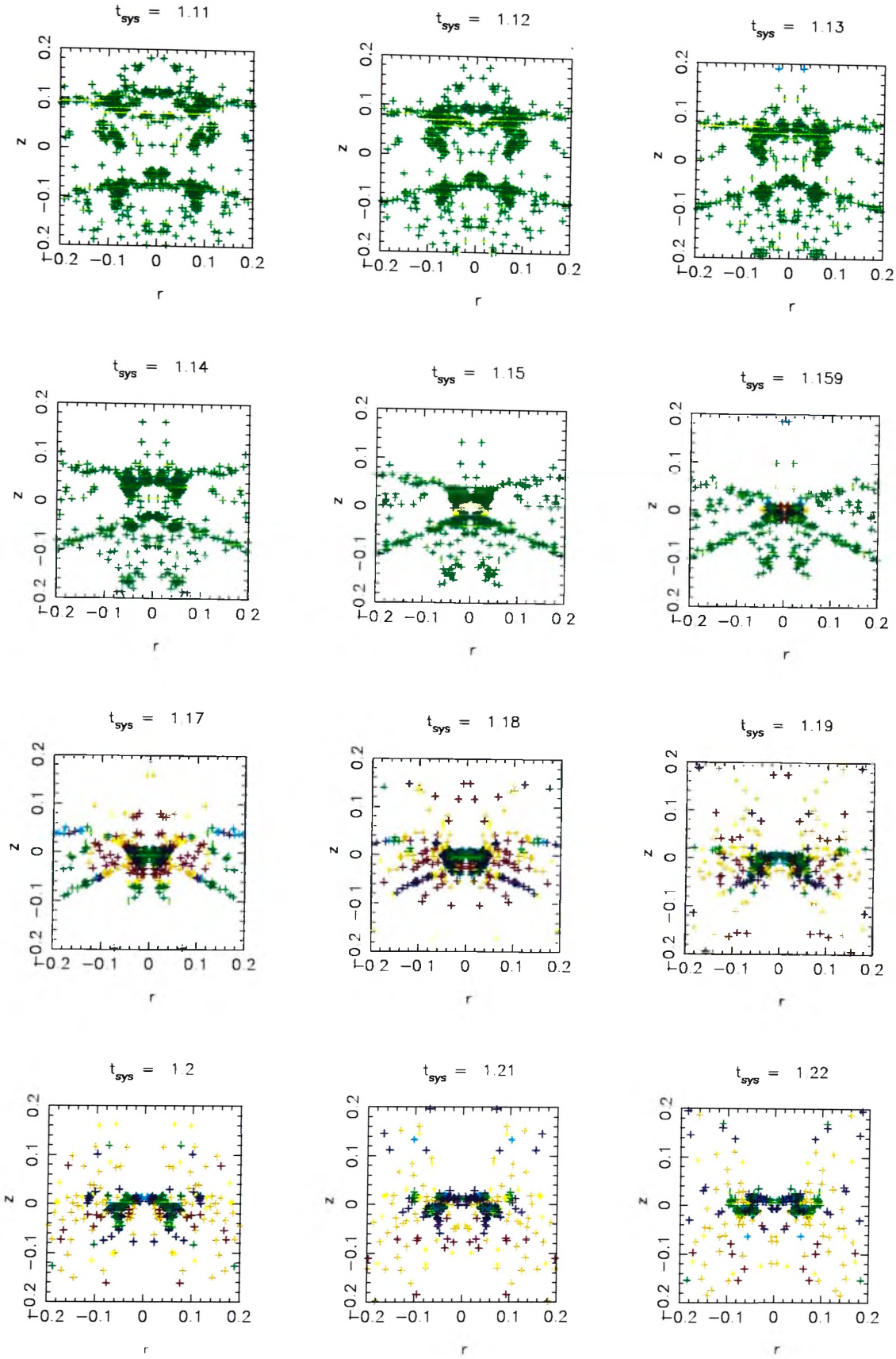

Figure 6.16: Temperature in the 2D model when cooling is reduced by the amount of viscous heating

solution for the in-shock cooling problem.

An explanation for the failure of this approach can be found when looking at the cooling function (Figure 5.1). The cooling function Λ, increases rapidly with temperature for lower temperatures. Reducing the cooling by the amount of viscous heating increases the temperature, thereby increasing the cooling rate over this range. This is confirmed by Figure 6.17 and Figure 6.18, which show the uncorrected cooling rate, $\varepsilon_{\rm rad}$, i.e. the rate at which the gas would cool if the cooling rate was not reduced by the amount of viscous heating, versus the viscous heating rate, $\varepsilon_{\rm vis}$. As in Figure 6.9, there is a strong correlation between $\varepsilon_{\rm vis}$ and $\varepsilon_{\rm rad}$, but the $\varepsilon_{\rm rad}$ is now about twice as large as $\varepsilon_{\rm vis}$, indicating that the implemented cooling rate $\varepsilon_{\rm rad*} = \varepsilon_{\rm rad} - \varepsilon_{\rm vis}$ equals the viscous heating rate. In other words, the gas adjusts so that it still radiates as much thermal energy away as it gains from viscous heating. The effect of the modification is to raise the gas temperature until the cooling becomes so efficient that viscous heating cannot increase the temperature any further. The method therefore does not resolve the in-shock cooling problem.

6.6 Turning Cooling Off in Shocks

Turning off cooling in shocks requires a decision as to whether a particle is in a shock or not. A very crude way of turning off cooling during shocks is turning all cooling off for a certain period. This is similar to the method tried by Maguire, of turning cooling on only after a certain time. The major disadvantage is that such a method does not distinguish between the particles. When cooling is turned off, it is turned off for all particles, with the result that physically realistic cooling

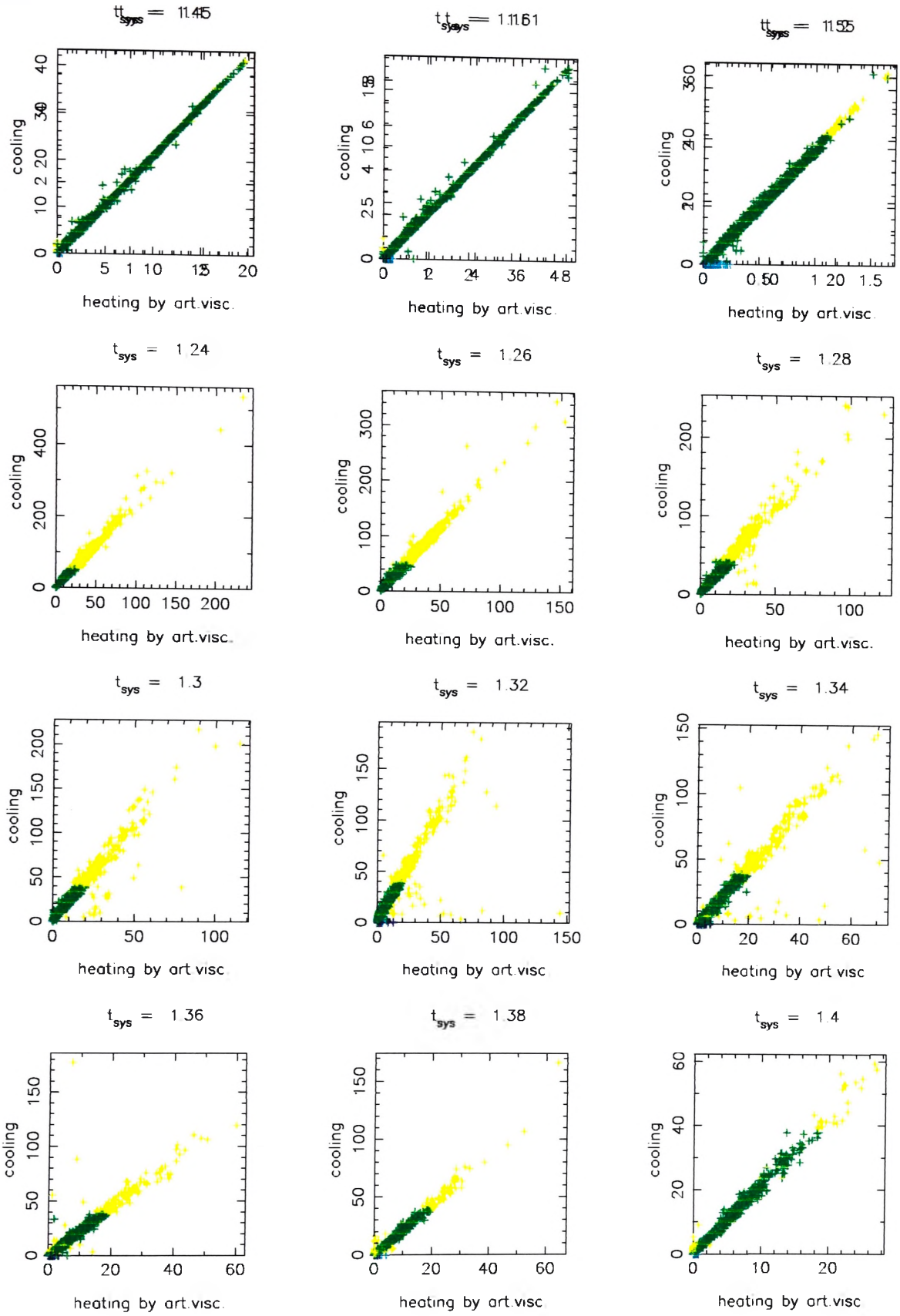

Figure 6.17: Viscous heating vs. the uncorrected radiative cooling rate when cooling is reduced by the amount of viscous heating, 3D model

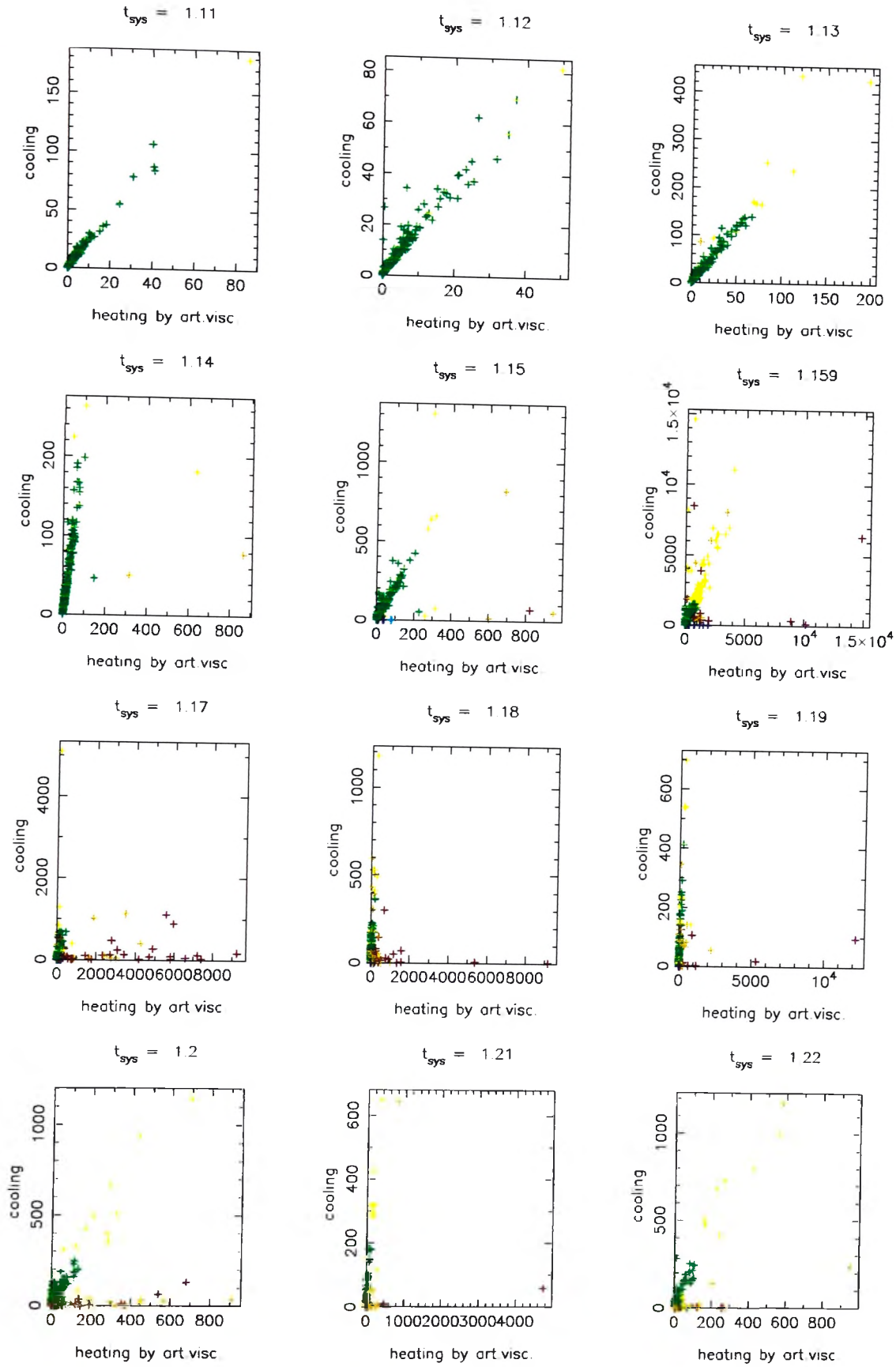

Figure 6.18: Viscous heating vs. the uncorrected radiative cooling rate when cooling is reduced by the amount of viscous heating, 2D model

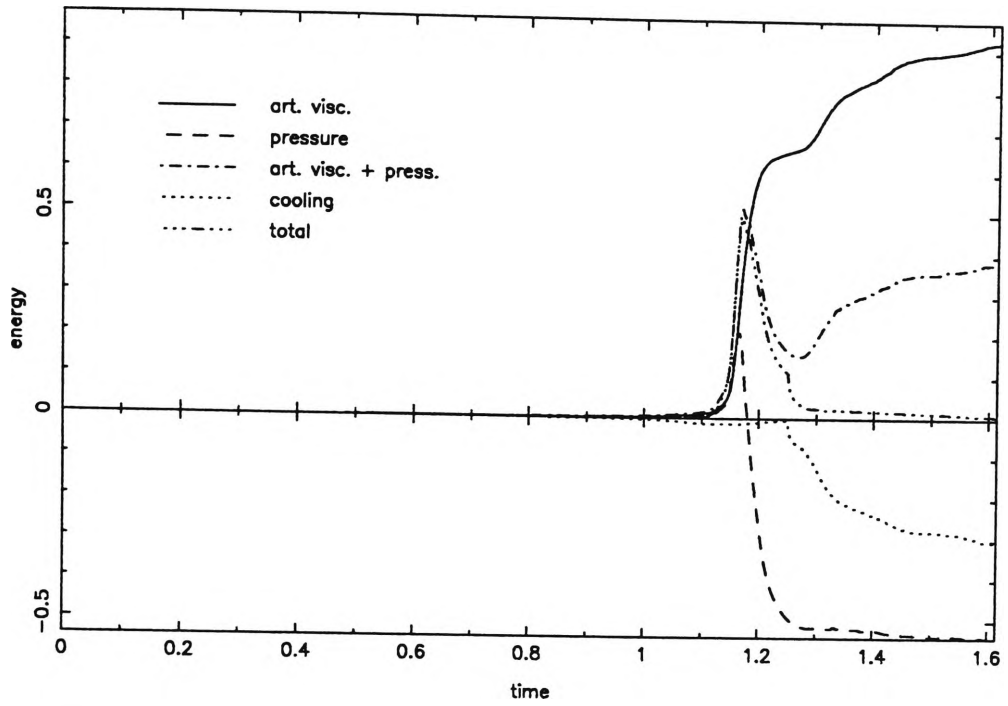

Figure 6.19: Heating and cooling in the 2D model if cooling is turned off between time 1.1 and 1.25

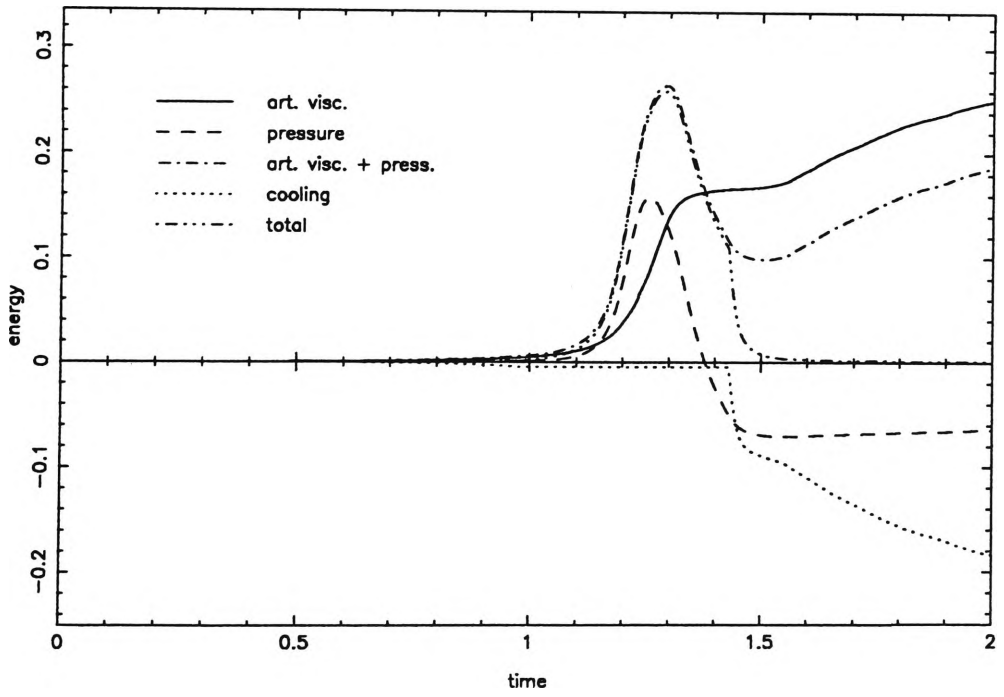

Figure 6.20: Heating and cooling in the 3D model if cooling is turned off between time 0.9 and 1.43

cannot take place. When cooling is switched on, particles undergoing shock heating will again be affected by the in-shock cooling problem. The method is easy to implement however, provided it is known when most shock heating occurs. Figure 6.19 and Figure 6.20 illustrate the problems with this method. When cooling is turned off, the temperature of the gas rises rapidly, followed by an adiabatic expansion. Once cooling is turned back on, the gas falls back in. It gains thermal energy by viscous heating and immediately radiates this away.

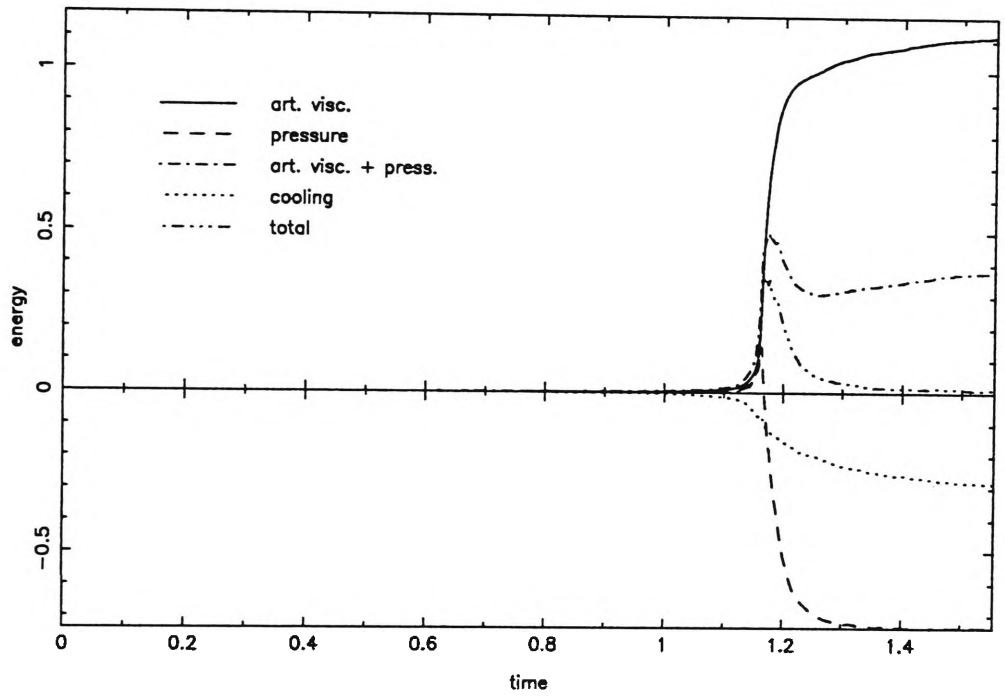

Figure 6.21: Heating and cooling in the 2D model if cooling is turned off for viscous heating rates > 5

A better way of deciding whether a particle is in a shock is to look at the viscous heating rate. This is not dimensionless and therefore does not fulfil the requirements mentioned above, but it gives an indication of the effect of turning cooling off in shocks. Simulations were run with cooling turned off for viscous heating rates > 5 in system units for both models. Comparison of Figures 6.1 and 6.3 with Figures 6.21 and 6.22 shows that in-shock cooling is clearly reduced and

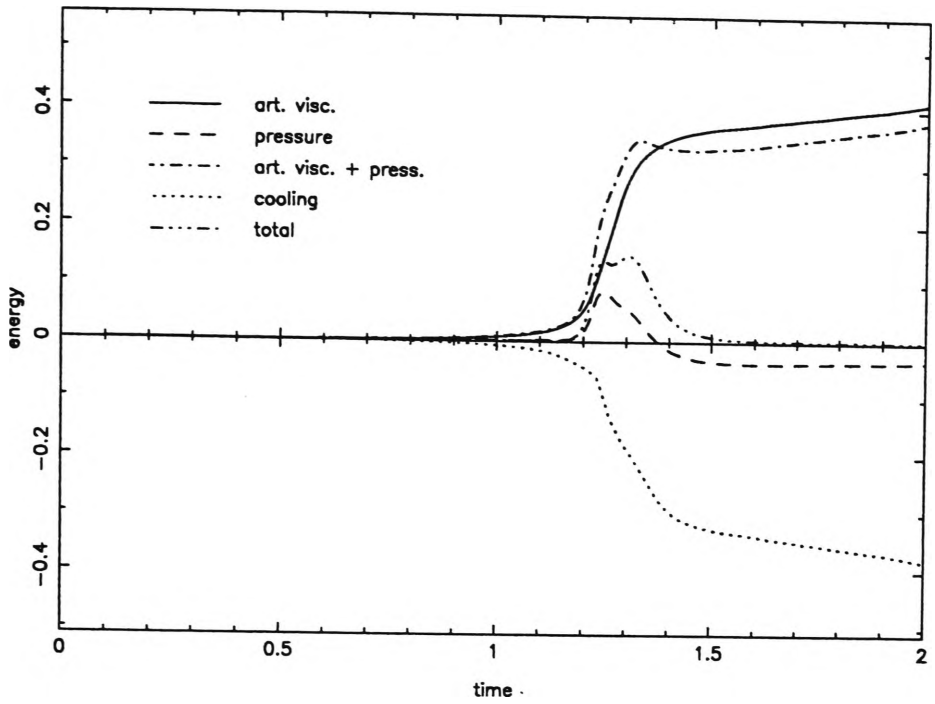

Figure 6.22: Heating and cooling in the 3D model if cooling is turned off for viscous heating rates > 5

that more hot gas is produced. This can also be seen by comparing Figures 6.13 and 6.15 with Figures 6.23 and 6.24. The effect is stronger in the 2D model than in the 3D model, as a larger proportion of the particles in the shock have viscous heating rates > 5 in the better resolved 2D model. In both models heating of the gas is followed by adiabatic expansion, reducing the amount of gas in the central region. Unfortunately, this criterion is not applicable over a wide dynamical range and must be changed to suit any particular situation.

6.6.1 Criteria to Identify Shocks

A good criterion to turn of cooling in shocks should fulfil the requirements outlined above. In particular, it should be dimensionless and select particles which are shocking and no others. In order to be independent of the system being simulated, it should be defined locally and, in order to be frame independent, it

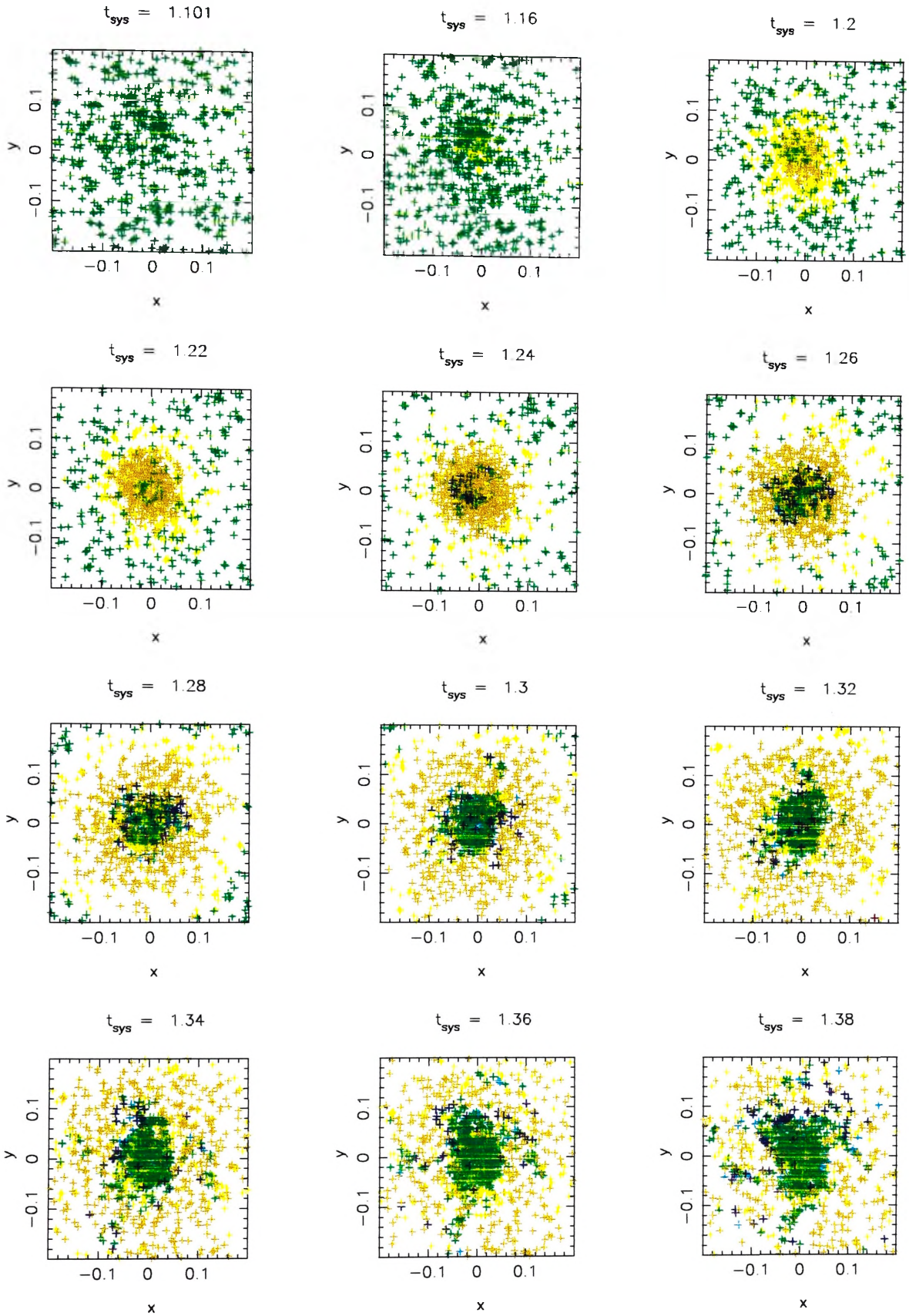

Figure 6.23: Temperatures in the 3D model with cooling turned off for viscous heating rates > 5

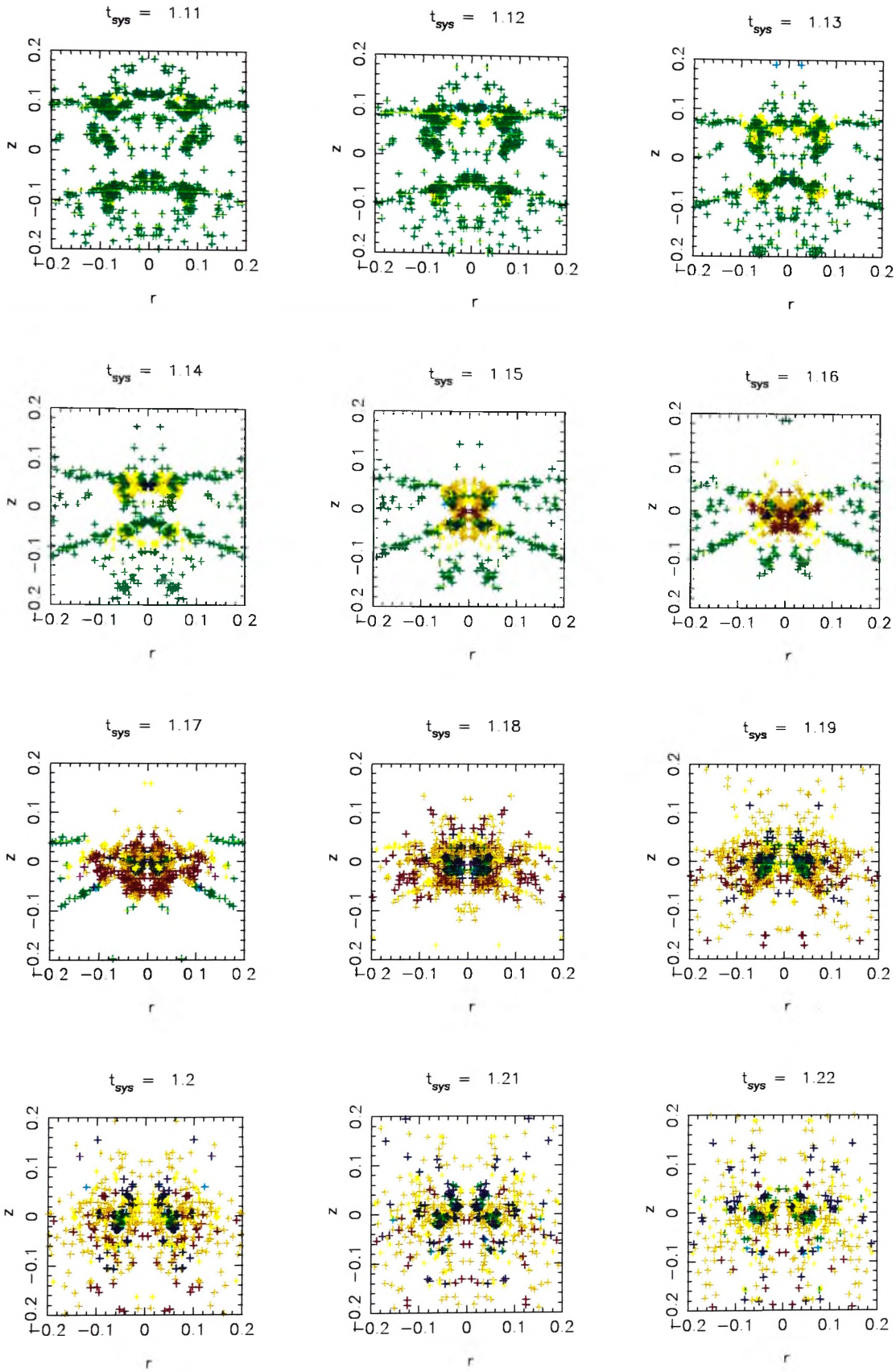

Figure 6.24: Temperatures in the 2D model with cooling turned off for viscous heating rates > 5

should not depend on the absolute velocity of the particles. Six possible criteria with these features were selected to be investigated.

Criterion 1 is based on the artificial viscosity parameter μ:

$$\frac{\mu_i}{c_i} > b \tag{6.18}$$

where b is to be determined later, c_i is the sound speed of particle i and

$$\mu_i = \sum_j m_j \frac{\mu_{ij}}{\rho_j} W_{ij} \tag{6.19}$$

where W_{ij} is the kernel and μ_{ij} is the artificial viscosity parameter μ_{ij} given by equation 5.27. μ_i is a measure of the speed of approach of particles about 1 kernel length away. Thus we expect μ_i/c_i to be of order unity or larger in a shock.

Criteria 2 to 4 are based on the viscous stress Π. In a shock the viscous stress is comparable to or larger than the pressure. Criterion 2 compares the viscous stress per unit mass with the average temperature (pressure per unit mass) in the neighbourhood of a particle.

$$\frac{\Pi_i}{p_i} > b \tag{6.20}$$

with Π_i

$$\Pi_i = \sum_j m_j \Pi_{ij} W_{ij} \tag{6.21}$$

where Π_{ij} is the artificial viscosity term given by equation 5.26 and p_i

$$p_i = \sum_j m_j \left(\frac{P_i}{\rho_i^2} + \frac{P_j}{\rho_j^2} \right) W_{ij} \tag{6.22}$$

Criterion 3 is similar to criterion 2, but defined to compress the range of the parameter:

$$\frac{\Pi_i}{p_i + \Pi_i} > b \tag{6.23}$$

Criterion 4 is similar to criterion 2 but takes the particle temperature instead of the average temperature:

$$\frac{\Pi_i}{T_i} > b \tag{6.24}$$

where T_i is the temperature of particle i in system units.

Criterion 5 is based on the viscous heating rate per unit mass. In order to obtain a dimensionless criterion, this has to be compared with another local quantity which has units of power/mass. This was achieved by using $T_i^{3/2}/h_i$:

$$\frac{\varepsilon_{\mathrm{vis,i}} h_i}{T_i^{3/2}} > b \tag{6.25}$$

where $\varepsilon_{\mathrm{vis,i}}$ is the viscous heating rate, given by equations 6.13 and 6.12 and h_i is the smoothing length for particle i.

Criterion 6 is similar to criterion 5 but instead of the viscous heating rate it uses the square root of the viscous heating rate times the cooling rate. While this criterion might seem a less valid way of selecting the shock, it might be more effective in selecting situations where the in-shock cooling is large. The criterion is

$$\frac{\sqrt{(\varepsilon_{\mathrm{vis,i}} \times \varepsilon_{\mathrm{cool,i}})} h_i}{T_i^{3/2}} > b \tag{6.26}$$

where $\varepsilon_{\mathrm{cool,i}}$ is the radiative cooling rate per unit mass $\rho k \Lambda$.

6.6.2 Evaluation of Criteria

The criteria of the previous section were all evaluated by comparing them with the viscous heating rate and the radiative cooling rate. Criteria which can be used to alleviate the in-shock cooling problem should distinguish particles which have high viscous heating rates and high cooling rates from particles with low viscous heating rates and low cooling rates. If a criterion is independent of the implementation of the model, the resolution and the number of spatial dimensions, it should be possible to apply the same value for b in both models. In order to evaluate the suitability of the different criteria, values of the quantities described above were plotted against the viscous heating rate at different times during the simulations. Here, plots are shown for a few selected times, with Figures 6.25 and 6.26 showing the quantities used for criteria 1-3 and 4-6 for the 2D model, and Figures 6.27 and 6.28 showing the quantities used for criteria 1-3 and 4-6 in the 3D model. The colours indicate the cooling rate, with the values given in section 6.3.

As can be seen, none of the quantities that were selected as criteria correlates well with the viscous heating or the cooling rate. In fact, the highest values for the quantities used as criteria are found for particles which have little viscous heating or radiative cooling, and the particles with the largest amount of viscous heating and radiative cooling have low values for the criteria. Even at an early stage of the collapse, when the gas is not yet shocking, the values of these quantities is already considerable. This finding applies to both models.

The explanation for the failure of these criteria is in the nature of the collapse and the wide dynamic range of temperatures involved. All of the quantities above

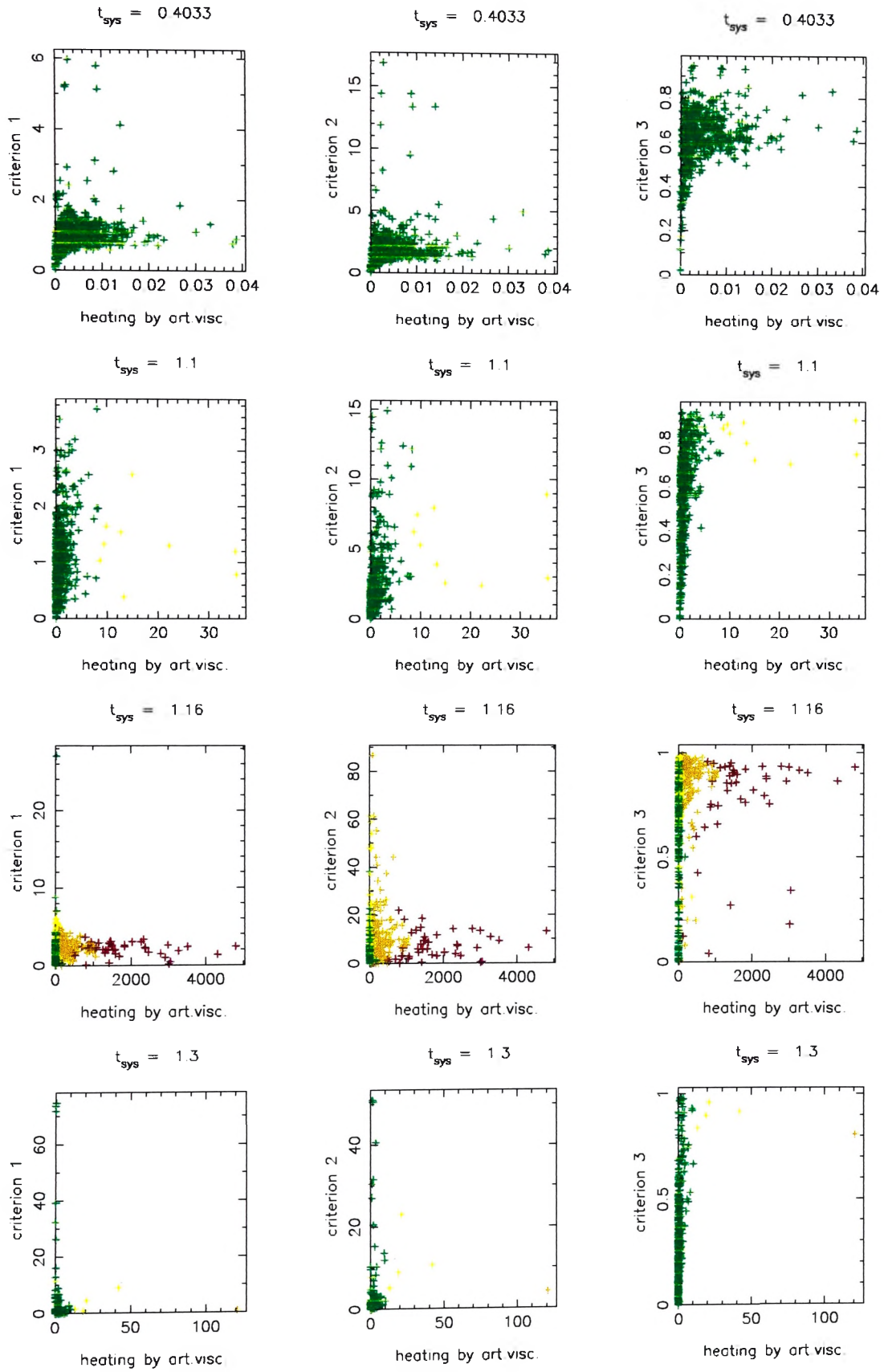

Figure 6.25: criteria 1,2 and 3 vs viscous heating for the 2D model

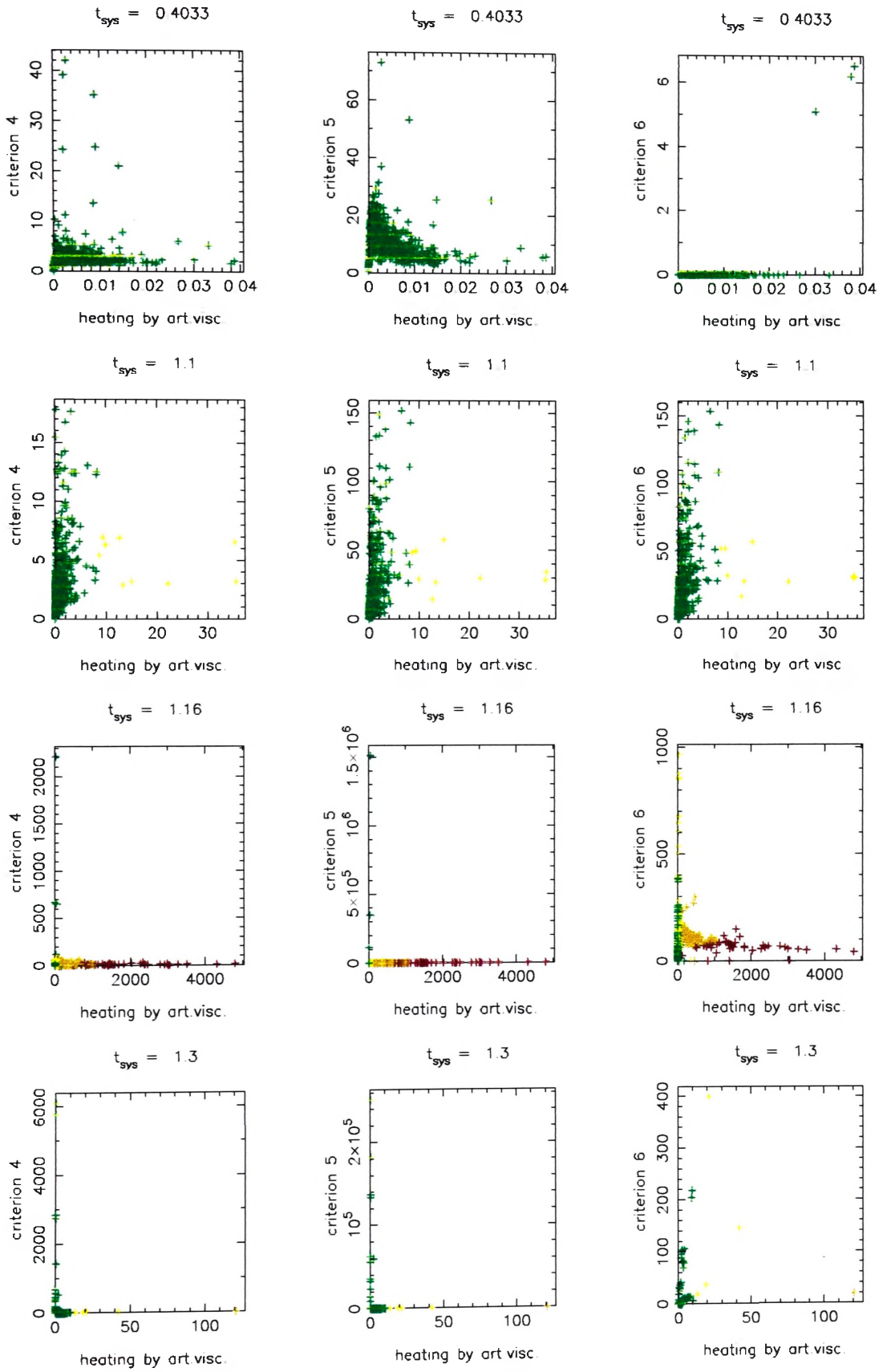

Figure 6.26: criteria 4,5 and 6 vs viscous heating for the 2D model

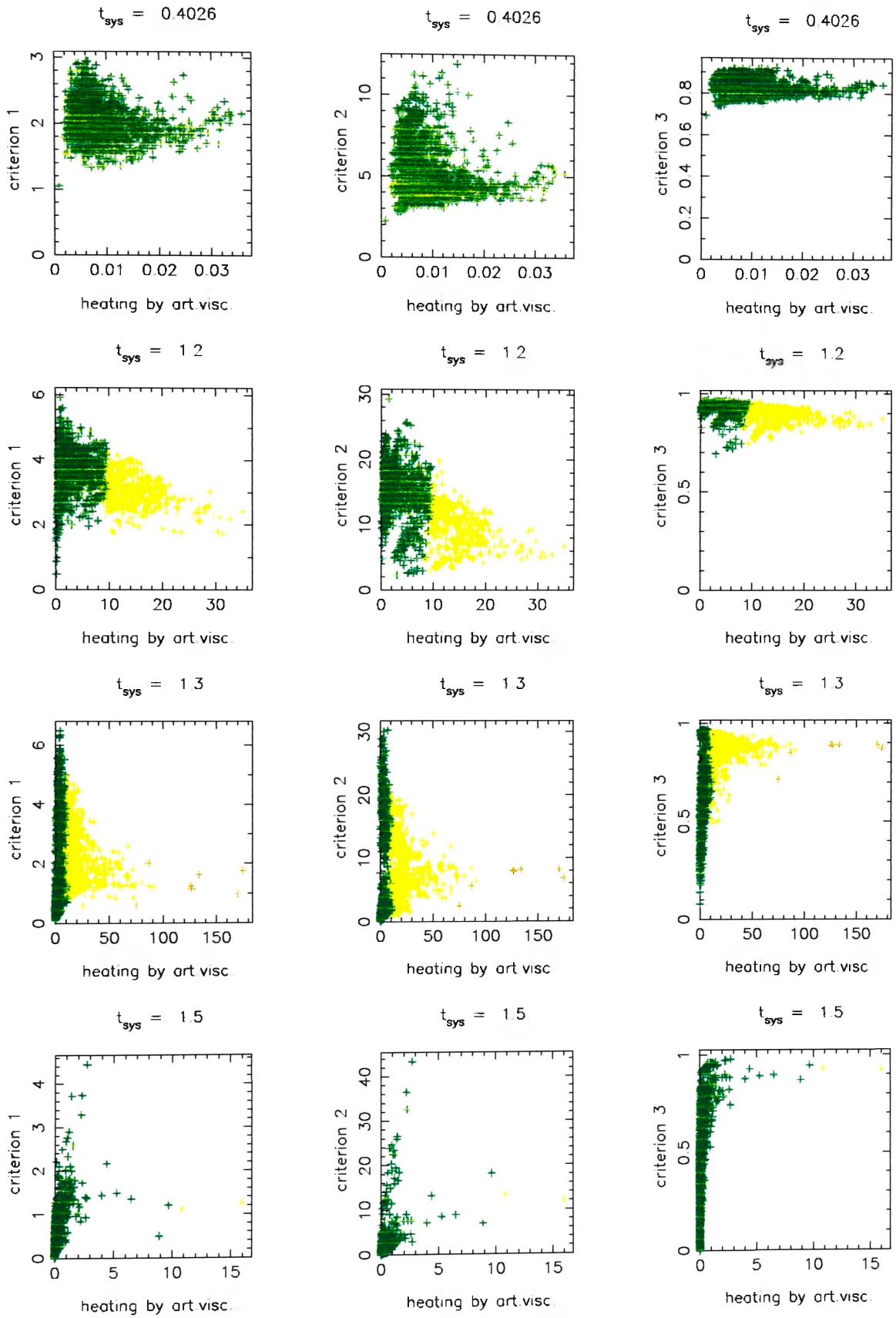

Figure 6.27: criteria 1,2 and 3 vs viscous heating for the 3D model

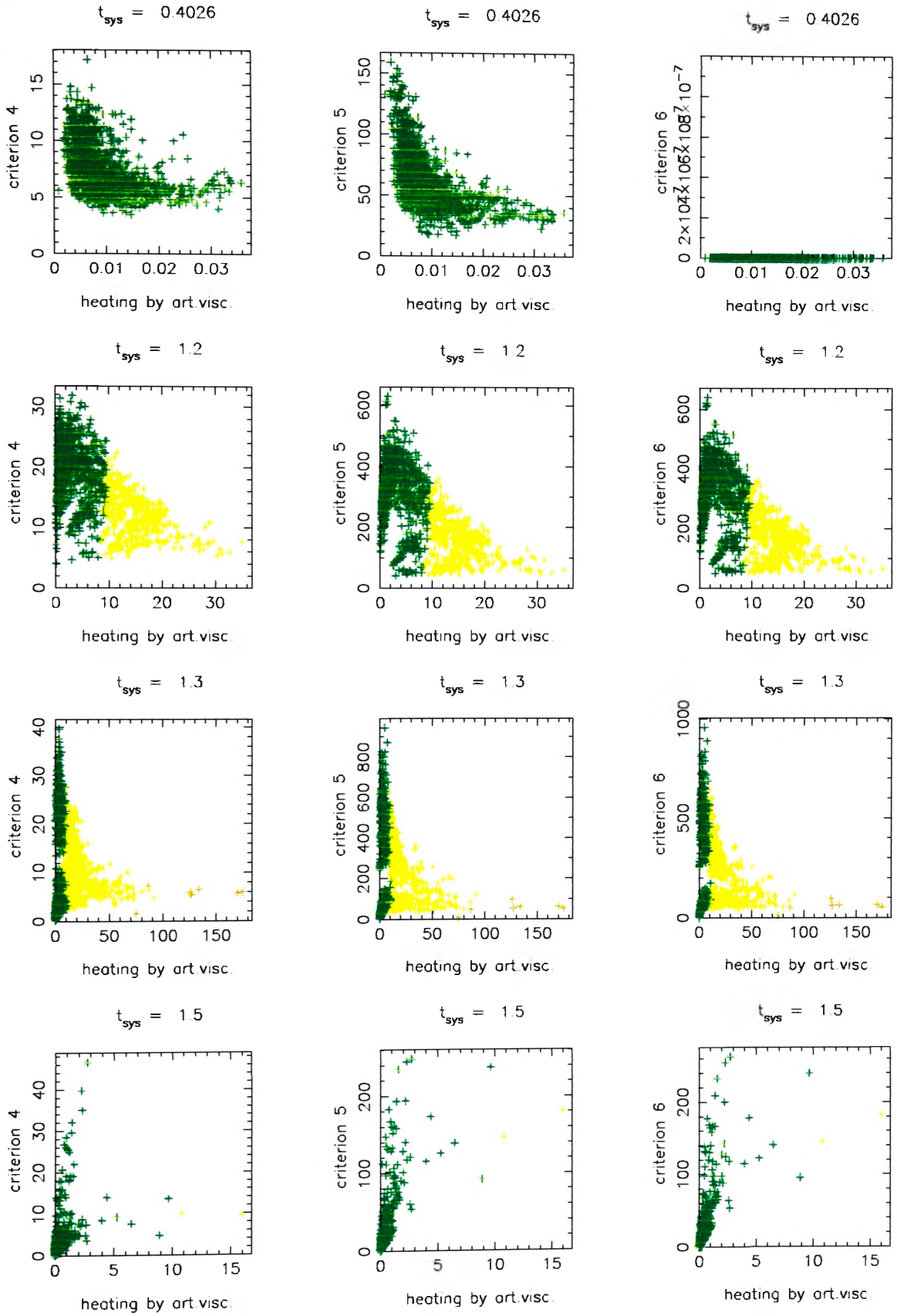

Figure 6.28: criteria 4,5 and 6 vs viscous heating for the 3D model

depend on the temperature, in such a way that, all other things being equal, larger values of the criteria occur for low temperatures. The only exception is the quantity used for criterion 6, which is 0 for temperatures below the cut off of the cooling function. Above this temperature, however, the quantity shows a similar temperature dependence to the other criteria.

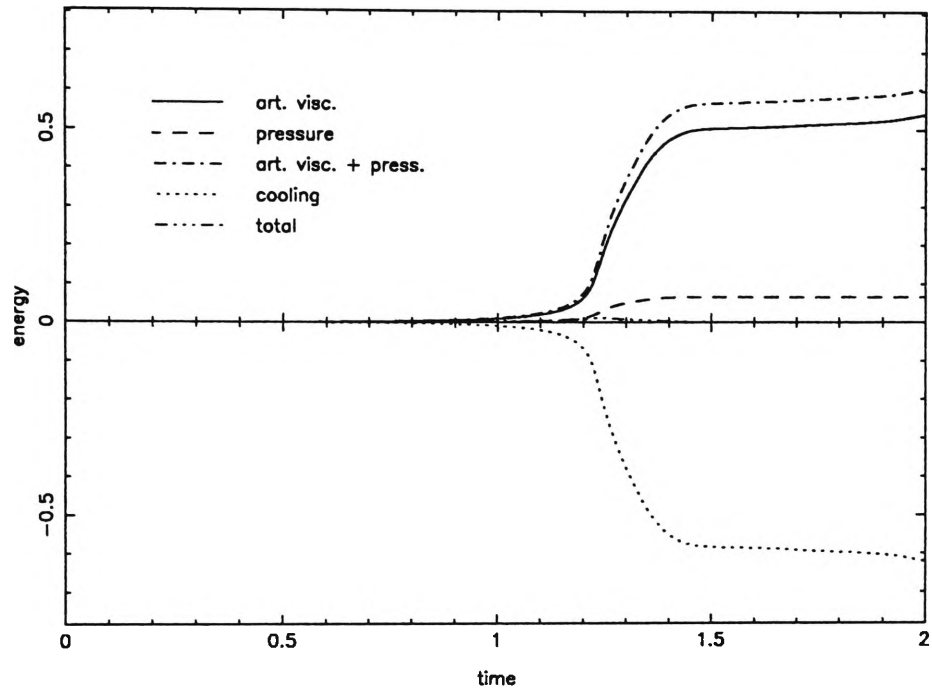

Figure 6.29: Heating and cooling in the 3D model if cooling is turned off for $\mu/c > 2$

The viscous heating rate depends on the rate of compression of the fluid. Since collapse involves compression, from the onset there is weak viscous heating but because of the low gas temperature, this appears significant under the criteria evaluated here.

So far, it has been assumed that the absolute viscous heating rate and cooling rate are what matter most, but is is possible that the criteria are still useful in identifying shocks under a wide dynamic range of conditions. To check this possibility, a number of the criteria were tested in simulations with different values

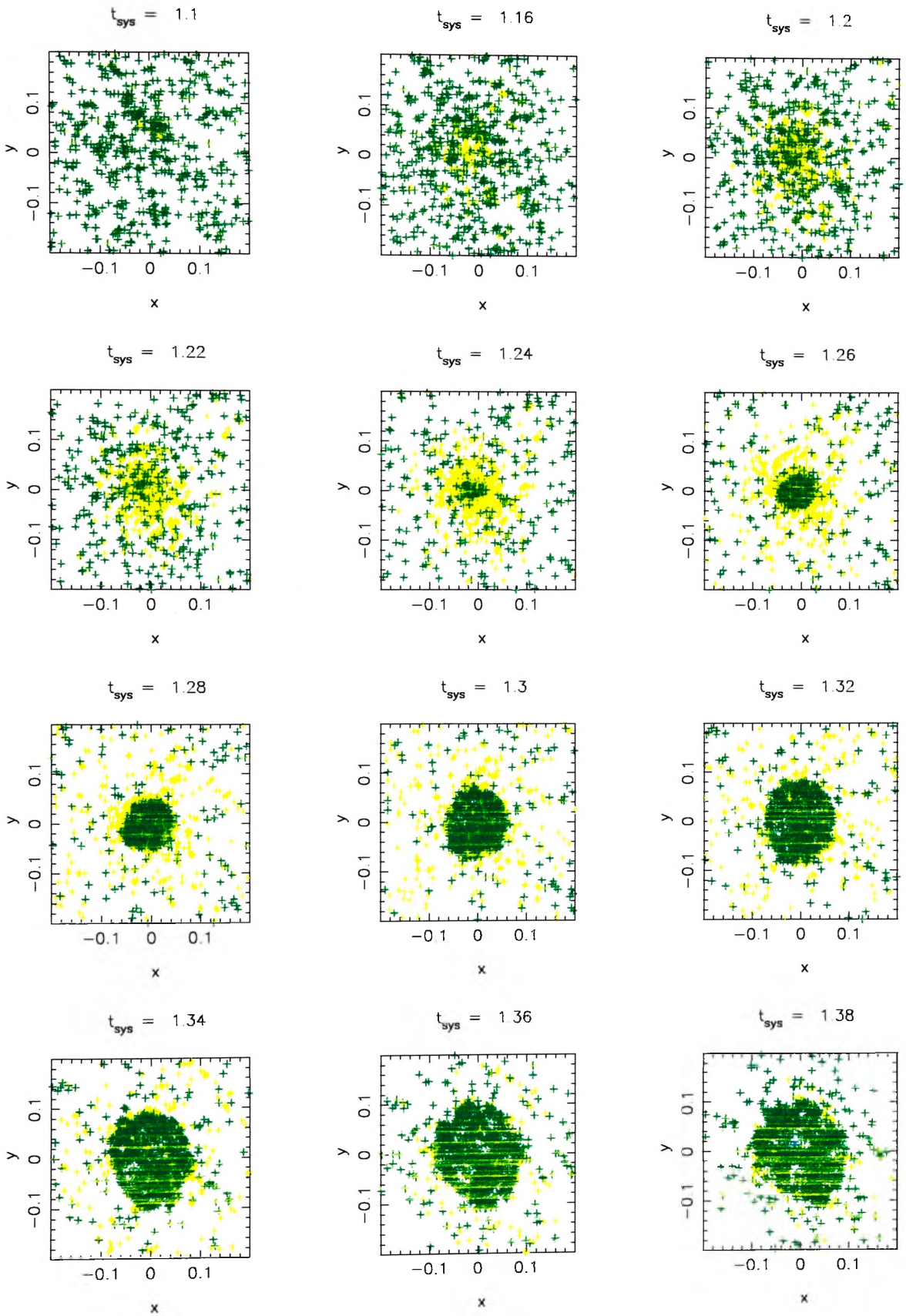

Figure 6.30: Temperatures in the 3D model with cooling turned off for $\mu/c > 2$

for parameter b. As anticipated, the results are unsatisfactory. For higher values of b, turning cooling off when the criteria were met did not result in a significant decrease in in-shock cooling while, for lower values of the criteria, cooling was turned off all the time. Here results for criterion 1 with $b = 2$ are shown (Figures 6.29 and 6.30). The modification only results in a slight increase in the thermal energy of the particles, even though cooling is turned off at some time for a large proportion of the particles.

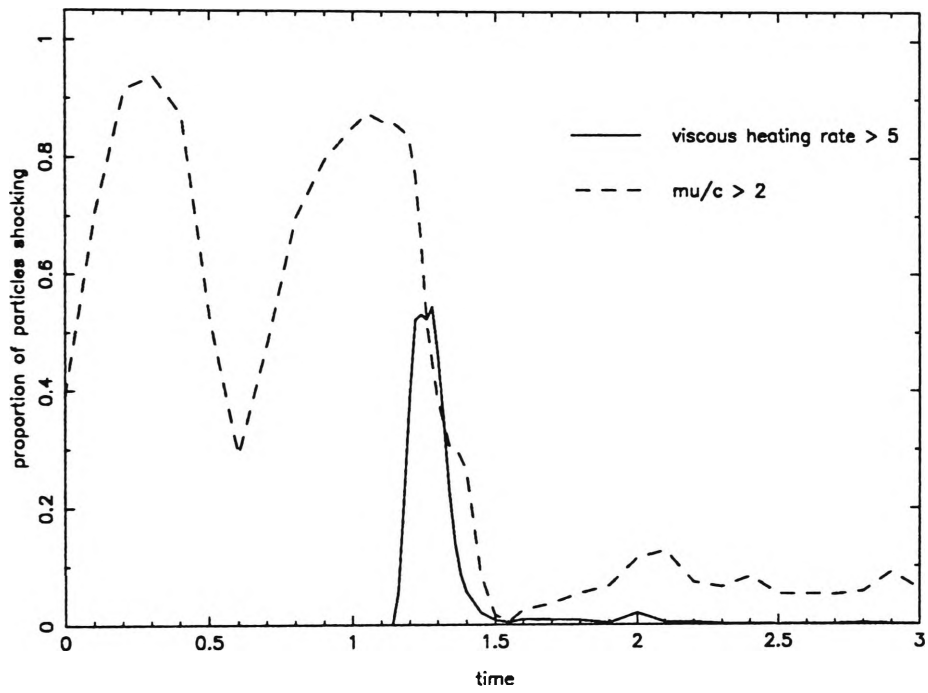

Figure 6.31: Proportion of particles shocking for $\mu/c > 2$ and for $\varepsilon_{vis} > 5$

Figure 6.31 shows the proportion of particles for which cooling was turned off in the simulations shown in Figure 6.23 and Figure 6.30. At almost any time during the simulation, cooling was turned off for a larger proportion of the particles when using $\mu/c > 2$ as criterion than when using $\varepsilon_{vis} > 5$. In spite of this, much more hot gas was produced in the latter simulation. Simulations for the other criteria show similar results. Clearly the criteria do a poor job of

selecting the right particles.

6.7 Discussion

The main result of this work is that none of the criteria proposed in section 6.6 for identifying particles in shocks performs well. The only criterion that had success in preventing in-shock cooling, $\varepsilon_{\mathrm{vis,i}} > 5$, is highly dependent on the particular problem and certainly not useful in flows involving a wide dynamic range of scales. Nevertheless, it does show that it is possible to ameliorate in-shock cooling. The failure of the criteria tried here leaves open the question of whether or not a dimensionless, universal criterion, that distinguishes between particles that are shocking and those that are not, exists. The criteria tried here, fail to distinguish between shocks and homologous collapse. The problem that a criterion has to be found to locate the shock in order to turn the cooling off is similar to the problem of that the shocks need to be located in order to switch artificial viscosity off when the gas is collapsing but not shocking. Ideally, artificial viscosity should be turned off and cooling should be switched on in these situations, while artificial viscosity should be switched on and radiative cooling should be switched off during a shock. Several authors have tried to find a method to switch artificial viscosity off for particles which are not shocking [84] [114] [102], but so far a dimensionless, situation independent criterion has not been found.

The fundamental problem with finding a suitable criterion seems to be in the requirement that it should be dimensionless. The quantities which change rapidly in a shock can all be expressed in terms of the relative velocities of the particles. This means that the quantity used for the criterion will be a function

of the relative velocities. In order to make the criterion dimensionless, another quantity must therefore be found, which also is a function of velocity or specific energy. In the collapses simulated here there are three forms of energy: kinetic energy, gravitational potential energy and thermal energy. The kinetic energy is determined by the absolute velocity of the particle and is therefore frame dependent. Even in situations where this would not be a large problem, as in the galaxy formation simulations performed here, the velocity would not be a good quantity to use as its value can be very small in situations where the gas is not shocking, such as at the start of the simulation or after the shock. This would result in large values for the quantity used as criterion while the gas is not in a shock. The gravitational potential energy is also not suitable for comparison as it is determined by the system being simulated and is therefore not really a local quantity. This leaves the thermal energy with the problem, described above, that low temperatures during the early part of the simulation result in large values of the quantity used as criterion. When the temperature rises significantly, as happen in a sufficiently strong shock, the value of the quantity goes down. The result is that if the thermal energy is used to create a dimensionless criterion, it does a poor job of selecting the right particles.

In view of this it is hard to be optimistic that it will be possible to develop widely applicable criteria for controlling in-shock cooling. Nevertheless there are measures that can control it to some extent. The competition between shock heating and radiative cooling plays a critical role in normal galaxy formation, so this problem must be solved if we are to make credible hydrodynamic simulations of galaxy formation.

Chapter 7

Conclusions

7.1 Testing NF's Assumption for Supernova Feedback in Low Mass Galaxies

The first part of this work concerns the effect of an assumption made by Nulsen and Fabian in their semi-analytical galaxy formation model. NF assume that in low mass systems star formation is regulated by supernova explosions which eject the remaining gas from the galaxy, thereby stopping further star formation. The assumption being tested here is that this happens when the energy released by supernova explosions exactly equals the binding energy of the remaining gas, i.e. the energy required to take the gas to infinity. The effect of this assumption is investigated in two different versions of the model, the model described in [100] which has a flat cosmology and the model described in [101] which has an open cosmology.

In the NF model the collapsed gas is separated into two parts, one part, the catastrophically cooled gas, which is cold immediately after the collapse and forms into stars and one part which will remain hot. A fraction of the catastrophically cooled gas will turn into the massive stars, above $\sim 8M_{\odot}$, which form type II supernovae. It is assumed that star formation proceeds until either all the catastrophically cooled gas is turned into stars or the amount of energy released by supernovae precisely equals the binding energy of the remaining gas, i.e. the remaining cold gas and any hot gas around it. The sensitivity of their model to this assumption is tested here.

For this work, the fraction of the catastrophically cooled gas that turns into stars, f_*, is no longer precisely that required to produce enough supernovae to unbind the remaining gas. Instead, the energy released by supernovae is taken

to be equal to the energy needed to unbind the rest of the gas and to give it an extra amount of specific energy, $\mu\sigma^2$. μ is a new parameter which can be either positive or negative.

In general the results of the simulations are found not to be sensitive to the assumption being tested The outcome that is most sensitive to this assumption is the ratio of elliptical to spiral galaxies, but this is known to be sensitive to other model parameters. Most other properties remain within error margins as the parameter μ is varied. The constancy of the total star formation with varying μ is remarkable as the star formation in small dwarf galaxies can change by up to 25%. The result of an increase in star formation during early collapses, however, is a decrease in the amount of remaining gas and an increase in its the energy content. This reduces the star formation resulting from subsequent collapses. Since the reduction in star formation during later collapses is directly related to the increase in star formation during early collapses, the total star formation is insensitive to the value of μ. This result, however, only applies to star formation in spheroids. Star formation in discs, which is not modeled here, might be sensitive to μ.

The reduction of the fraction of elliptical galaxies with increasing μ is its most significant effect. In the NF model a disc galaxy forms when the last of the hot gas cools. The disc formation rate, found by taking the whole of each hot halo which finishes cooling to be turned into a disc, shows a complicated dependence on μ. There are several factors which influence the disc formation rate and are affected by changing μ. Increasing μ results in an earlier transition from dwarf to normal galaxies. The temperature in these galaxies is low, increasing the

chance that all the remaining gas cools before the next collapse. However, the time between collapses for these galaxies is short, and this counteracts the first effect. The reduced gas fraction and the increased energy content with increasing values of μ, as mentioned above, also reduce the chance that the gas can cool before the next collapse. The competition between these effects is complicated by further factors, such as the temperature and metallicity dependence of the cooling function. Which of these effect dominates depends subtly on the values of μ and the redshift.

The effect of modifying the collapse model is qualitatively similar in high and low density cosmological models although there are quantitative differences. In the low density model, most effects are smaller and the transition from increased to decreased star formation rates with increasing values of μ takes place at a lower redshift. The reduction in the elliptical fraction with increasing values of μ is similar in the two models but this fraction is higher in the low density model for all values of μ.

The objective of this work was to determine if the NF model is sensitive to one of its weakest assumptions. The results here show that it is not sensitive to modest changes in that assumption. The model outcome that is most sensitive to the assumption, galaxy morphology, is more sensitive to several other model parameters, so that it is not made significantly less certain by this assumption. It can be concluded that the assumption made by Nulsen and Fabian, that total star formation proceeds until either all the catastrophically cooled gas is turned into stars or the amount of energy released by supernovae is just enough to unbind the rest of the gas, has a relatively minor influence on the results of their model.

7.2 The In-shock Cooling Problem

The second part of this work involves an attempt to find a solution for the in-shock cooling problem. The in-shock cooling problem is the result of limited spatial resolution in numerical simulations. Simulated shocks can be more than four orders of magnitude thicker. The gas therefore needs much more time to cross a simulated shock. While in reality the shock-crossing time is so short that cooling can be ignored, the gas can cool considerably in numerical shocks shocks. This prevents it from ever reaching the correct post-shock temperature. The problem was investigated using two models for the formation of a disc galaxy, a 2 dimensional axisymmetric SPH model and a 3 dimensional SPH model.

It was demonstrated that the lack of hot gas produced in galaxy formation simulations is indeed the result of in-shock cooling. In both models most thermal energy obtained by viscous heating is radiated away almost simultaneously. The possibility that this is the result of radiative cooling of particles which have already passed through the shock, while other particles, which have not yet passed through the shock are still heating was excluded by comparing the viscous heating rate per unit mass with the radiative cooling rate per unit mass for individual particles. It was shown that the viscous heating rate is strongly correlated with the radiative cooling rate.

Two methods were tried to ameliorate the in-shock cooling problem, reducing radiative cooling by the amount of viscous heating and turning off cooling in shocks. The first method did not produce satisfactory results. This is caused by the very efficient cooling at lower temperature and by the fact that for these temperatures the cooling function rises rapidly with temperature. Reducing the

amount of radiative cooling by the amount of viscous heating results in a slightly higher temperature, but at this temperature cooling is more efficient. Once a temperature is reached where the uncorrected cooling rate is twice the viscous heating rate, all thermal energy gained by viscous heating is radiated away and the temperature does not increase further. This is what happened in the simulations, although the modification produced slightly better results for the 2D than for the 3D model.

The second method, turning cooling off in shocks requires a criterion to decide whether a particle is in a shock or not. Applying such a criterion should result in reduced in-shock cooling, but should not interfere with other aspects of the code and in particular should not reduce cooling for particles which are not in a shock. The criterion should not result in a large number of extra calculations or, in order to be universally applicable, should be dimensionless and not depend on what is simulated or on the details of the implementation. The criteria were all evaluated by comparing them with the viscous heating rate and the radiative cooling rate. Criteria which can be used to alleviate the in-shock cooling problem should distinguish particles which have high viscous heating rates and high cooling rates from particles with low viscous heating rates and low cooling rates.

Six dimensionless criteria were selected and their effectiveness in preventing in-shock cooling was evaluated. None of these criteria performed well. They did not select particles which were shocking and were therefore not successful in preventing in-shock cooling. It was possible, however, to significantly reduce in-shock cooling by turning off cooling for particles which have a viscous heating rate > 5 in system units. Unfortunately, this criterion is highly dependent on

the particular problem and certainly not useful in flows involving a wide dynamic range of scales. Nevertheless, it does show that it is possible to ameliorate in-shock cooling. The difficulty of finding a dimensionless criterion to locate the shock was noted before by authors who wanted to switch off artificial viscosity for particles which are not shocking [84] [114] [102]. The criteria do not distinguish between homologous collapse and shocks.

The fundamental problem with finding a suitable criterion seems to be in the requirement that it should be dimensionless. This means that two quantities of the same dimensions have to be compared. The quantities which change rapidly in a shock can all be expressed in terms of the relative velocities of the particles, so one of the quantities used for the criterion will therefore be a function of the relative velocities. In order to make the criterion dimensionless, another quantity must be found, which also is a function of velocity or specific energy. The possibilities are limited. In the collapses simulated here there are only three forms of energy: kinetic energy, gravitational potential energy and thermal energy. The kinetic energy is determined by the absolute velocity of the particle and is therefore dependent on the frame of reference. Even in situations where this would not be a large problem, like the galaxy formation simulations presented here, the velocity would not be a good quantity to use. Its value can be very small in situations where the gas is not shocking, such as at the start of the simulation or after the shock, which would result in selection of the wrong particles. The gravitational potential energy is not suitable as it is determined by the system being simulated and is therefore not really a local quantity. This leaves the thermal energy with the problem that low temperatures during the early part of

the simulation result in a large number of particles which are supposedly shocking. When the temperature rises significantly, as happen in a sufficiently strong shock, this number decreases. The result is that if the thermal energy is used to create a dimensionless criterion, it does a poor job of selecting the right particles.

In view of this it is hard to be optimistic that it will be possible to develop widely applicable criteria for controlling in-shock cooling. Nevertheless there are measures that can control it to some extent. The competition between shock heating and radiative cooling plays a critical role in normal galaxy formation, so this problem must be solved if we are to make credible hydrodynamic simulations of galaxy formation.

Bibliography

[1] M. Abramowitz and I.A. Stegun. *Handbook of Mathematical Functions.* Dover, New York, 1965.

[2] C. Alcock. MACHOS and Other Dark Subjects: What has been learned using gravitational lensing. *American Astronomical Society,* 195:124.01, 1999.

[3] C. Alcock, R.A. Allsman, D.R. Alves, T.S. Axelrod, A.C. Becker, D.P. Bennett, K.H. Cook, N. Dalal, A.J. Drake, K.C. Freeman, M. Geha, K. Griest, M.J. Lehner, S.L. Marshall, D. Minniti, C.A. Nelson, B.A. Peterson, P. Popowski, M.R. Pratt, P.J. Quinn, C.W. Stubbs, W. Sutherland, A.B. Tomaney, T. Vandehei, and D. Welch. The MACHO Project: Microlensing Results from 5.7 Years of LMC Observations. *ApJ Submitted, astro-ph,* 0001272:1–53, 2000.

[4] P. Anninos, M.L. Norman, and D.A. Clarke. Hierarchical Numerical Cosmology with Hydrodynamics: Methods and Code Tests. *ApJ,* 436:11–22, 1994.

[5] N.A. Bahcall. Large-Scale Structure in the Universe Indicated by Galaxy Clusters. *ARA&A,* 26:631–686, 1988.

[6] D.S. Balsara. *J.Comp.Phys.*, 121:357–372, 1995.

[7] J. Barnes and P.H Hut. *Nature*, 324:446–49, 1986.

[8] M.R. Bate and A. Burkert. Resolution Requirements for Smoothed Particle Hydrodynamics Calculations with Self-gravity. *MNRAS*, 288:1060–1072, 1997.

[9] A.C. Becker and et al. The MACHO Project: Microlensing Results from 5.7 Years of LMC Observation. *American Astronomical' Society*, 195:4802+, 1999.

[10] A.J. Benson, S. Cole, C.S. Frenk, C.M. Baugh, and C.G. Lacey. The Nature of Galaxy Bias and Clustering. *MNRAS*, 311-4:793–808, 2000.

[11] A.J. Benson, F.R. Frenk, C.S. Frenk, C.M. Baugh, and A. Jenkins. A Comparison of Semi-Analytical and Smoothed Particle Hydrodynamics Galaxy Formation. *astro-ph*, 9912220:1–13, 1999.

[12] W. Benz, R.L. Bowers, A.G.W. Cameron, and W.H. Press. Dynamic Mass Exchange in Doubly Degenerate Binaries. *ApJ*, 348:647–667, 1990.

[13] E. Bertschinger. Simulations of Structure Formation in the Universe. *ARA&A*, 36:599–654, 1998.

[14] B. Binggeli, A. Sandage, and G.A. Tammann. The Luminosity Function of Galaxies. *ARA&A*, 26:509–560, 1988.

[15] J. Binney, J. Kormendy, and S.D.M. White. *Morphology and Dynamics of Galaxies.* 12th Advanced Course, Swiss Society of Astronomy and Astrophysics, Geneva Observatory, CH-1290 Sauverny, Switserland, 1982.

[16] J. Binney and S. Tremaine. *Galactic Dynamics.* Princeton University Press, 41 William Street, Princeton, NJ 08540, USA, 1987.

[17] H. Böhringer and G. Hensler. Metallicity-dependence of Radiative Cooling in Optically Thin, Hot Plasmas. *A&A*, 215:147–149, 1989.

[18] J.R. Bond, S. Cole, G. Efstathiou, and N. Kaiser. Excursion Set Mass Functions for Hierarchical Gaussian Fluctuations. *ApJ*, 379:440–460, 1991.

[19] R.G. Bower. The Evoluation of Groups of Galaxies in the Press-Schechter Formation. *MNRAS*, 248:332–352, 1991.

[20] G.L. Bryan and M.L. Norman. Simulating X-ray Clusters with Adaptive Mesh Refinement. *AAS*, 187:1–3, 1995.

[21] E. F. Bunn and M. White. The 4 Year COBE Normalization and Large-Scale Structure. *ApJ*, 480:6+, 1997.

[22] S. Burles and D. Tytler. The Deuterium Abundance toward Q1937-1009. *ApJ*, 499:699+, 1998.

[23] S. Burles and D. Tytler. The Deuterium Abundance toward QSO 1009+2956. *ApJ*, 507:732–744, 1998.

[24] R. Capuzzo-Dolcetta and R. Di Lisio. A Criterion for the Choice of the Interpolation Kernel in Smoothed Particle Hydrodynamics. *astro-ph*, 9907089:1–12, 1999.

[25] B. Carr. Baryonic Dark Matter. *ARA&A*, 32:531–590, 1994.

[26] G. Carraro, C. Lia, and C. Chiosi. Galaxy Formation and Evolution - I. The Padua TREE-SPGH Code. *MNRAS*, 297:1021–1040, 1998.

[27] R. Cen. A hydrodynamic approach to cosmology - Methodology. *ApJS*, 78:341–364, 1992.

[28] R.Y. Cen, J.P. Ostriker, A. Jameson, and F. Liu. The universe in a box - Thermal effects in the standard cold dark matter scenario. *ApJ*, 362:L41–L45, 1990.

[29] R.A. Chevalier. The Interaction of Supernovae with the Interstallar Medium. *ARA&A*, 15:175–196, 1977.

[30] R.A. Chevalier. Emission from Circumstellar Interaction in Normal Type II Supernovea. *ApJ*, 420:268–285, 1994.

[31] S. Cole, A. Aragon-Salamanca, C.S. Frenk, J.F. Navarro, and S.E. Zepf. A Recipe for Galaxy Formation. *MNRAS*, 271:781–806, 1994.

[32] S. Cole and N. Kaiser. Sunyaev-Zel'dovich Fluctuations in the Cold Dark Matter Scenario. *MNRAS*, 233:637–648, 1988.

[33] H.M.P. Couchman. Mesh-Refined P3M: A Fast Adaptive N-Body Algorithm. *ApJ*, 368:L23–L26, 1991.

[34] H.M.P. Couchman, P.A. Thomas, and F.R. Pearce. Hydra: An Adaptive-Mesh Implementation of P3M-SPH. *ApJ*, 452:797–813, 1995.

[35] R. Dave, J. Dubinski, and L. Hernquist. Parallel TreeSPH. *New Astronomy*, 2-3:277–297, 1997.

[36] S. de Grandi, L. Guzzo, H. Böhhringer, S. Molendi, G. Chincarini, C. Collins, R. Cruddace, D. Neumann, S. Schindler, P. Schuecker, and W. Voges. The X-ray Luminosity function of bright galaxy clusters in the local universe. *ApJL*, 513:L17–L20, 1999.

[37] A. Dekel and J.P Ostriker. *Formation of structure in the Universe.* Cambridge University Press, 1999.

[38] A. Dekel and J. Silk. The Origin of Dwarf Galaxies, Cold Matter, and Biased Galaxy Formation. *ApJ*, 303:39–55, 1986.

[39] I.P dell'Antonioi, G.D. Bothun, and M.J. Geller. Peculiar Velocities for Galaxies in the Great Wall I: The Data. *ApJ*, 112:1759+, 1996.

[40] I.P dell'Antonioi, G.D. Bothun, and M.J. Geller. Peculiar Velocities for Galaxies in the Great Wall II: Analysis. *ApJ*, 112:1780+, 1996.

[41] R. Dominguez-Tenreiro, P.B. Tissera, A, and Saiz. Disk Formulation in Hierarchical Hydrodynamical Simulations: A Way Out of the Angular Momentum Catastrophe. *The Astrophysical Journal*, 508-2:L123–L127, 1998.

[42] B.T. Draine and C.F. McKee. Theory of Interstellar Shocks. *ARA&A*, 31:373–432, 1993.

[43] A.E. Evrard. Beyond N-Body: 3D Cosmological Gas Dynamics. *MNRAS*, 235:911–934, 1988.

[44] A.E. Evrard, F.J. Summers, and M. Davis. Two-Fluid Simulations of Galaxy Formation. *ApJ*, 422:11–36, 1994.

[45] A.C. Fabian. Cooling flows in clusters of galaxies. *ARA&A*, 32:277–318, 1994.

[46] S.M. Fall and G. Efstathiou. Formation and Rotation of Disk Galaxies with Haloes. *MNRAS*, 193:189–206, 1980.

[47] W. Forman and C. Jones. X-ray Imaging Observations of Clusters of Galaxies. *ARA&A*, 20:547–585, 1982.

[48] C.S. Frenk, A.E. Evrard, S.D.M. White, and F.J. Summers. Galaxy Dynamics in Clusters. *ApJ*, 472:460–484, 1996.

[49] C.S. Frenk, S.D.M. White, P. Bode, J.R. Bond, R. Cen, H.M.P. Couchman, A.E. Evrard, N. Gneding, A. Jenkins, A.M. Khokhloand A. Klypin, J.F. Navarro, M.L. Norman, J.P Ostriker, J.M. Owen, F.R. Pearce, U.L. Pen, M. Steinmetz, P.A. Thomas, J.V. Villumsen, J.W. Wadsley, M.S. Warren, G. Xu, and G. Yepes. The Santa Barbara Cluster Comparison Project: A Comparison of Cosmological Hydrodynamics Solutions. *ApJ*, 525:554–582, 1999.

[50] S. Ghigna, B. Moore, F. Governato, G. Lake, T. Quinn, and J. Stadel. Density Profiles and Substructure of Dark Matter Halos: "Converging" Results at Ultra-high Numerical Resolution. *astro-ph*, 9910166:1–19, 1999.

[51] R.A. Gingold and J.J. Monaghan. Smoothed particle hydrodynamics - Theory and application to non-spherical stars. *MNRAS*, 181:375–389, 1977.

[52] R.A. Gingold and J.J. Monaghan. Kernel Estimates as a Basis for General Particle Methods in Hydrodynamics. *Journal of Computational Physics*, 46:429–453, 1982.

[53] M. Girardi, S. Borgani, G. Giuricin, F. Mardirossian, and M. Mezzetti. The Observational Mass Function of Nearby Galaxy Clusters. *ApJ*, 506:45–52, 1998.

[54] N.Y. Gnedin. Galaxy Formation in a CDM + A Universe. I. Properties of Gas and Galaxies. *ApJ*, 456:1–33, 1996.

[55] K. M. Gorski, B. Ratra, R. Stompor, N. Sugiyama, and A. J. Banday. COBE-DMR-normalized Open Cold Dark Matter Cosmogonies. *ApJS*, 114:1+, 1998.

[56] L. Grego, J.E. Carlstrom, M.K. Joy, E.D. Reese, G.P. Holder, S. Patel, A.R. Cooray, and W.L. Holzapfel. The Sunyaev-Zel'dovich Effect in Abell 370. *astro-ph*, 0003085:1–17, 2000.

[57] J.E. Gunn and J.R. Gott. On the Infall of Matter into Clusters of Galaxies and Some Effects on Their Evolution. *ApJ*, 176:1–19, 1972.

[58] J.P. Henry and K.A. Arnaud. A Measurement of the Mass Fluctuation Spectrum from the Cluster X-ray Temperature Function. *ApJ*, 372:410–418, 1991.

[59] L. Hernquist and N. Katz. TREESPH: A Unification of SPH with the Hierarchical Tree Method. *AJS*, 70:419–446, 1989.

[60] L. Herquist. Some Cautionary Remarks about Smoothed Particle Hydrodynamics. *ApJ*, 404:717–722, 1993.

[61] J.S. Heyl, S. Cole, C.S. Frenk, and J.F. Navarro. Galaxy Formation in a Variety of Hierarchical Models. *MNRAS*, 274:755–768, 1995.

[62] R.W. Hockney and J.W. Eastwood. *Computer Simulations using Particles.* Hilger, Bristol, 1988.

[63] D. J. Horner, R. F. Mushotzky, and C. Scharf. Scaling laws and the mass and temperature functions of galaxy clusters. In *American Astronomical Society Meeting*, volume 194, pages 8903+, 1999.

[64] P. Hut. Smooth Particle Hydrodynamics: Models, Applications, and Enabling Technologies. *astro-ph*, 9710212:1–12, 1997.

[65] R.M. Hutchings and P.A. Thomas. In-shock Cooling in Numerical Simulations. *astro-ph*, 9903320:1+, 1999.

[66] H. Kang, J.P. Ostriker, R. Cen, D. Ryu, L. Hernquist, A.E. Evrard, G.L. Bryan, and M.L. Norman. A Comparison of Cosmological Hydrodynamic Codes. *ApJ*, 430:83–100, 1994.

[67] N. Katz. Dissipational Galaxy Formation. II. Effects of Star Formation. *ApJ*, 391:502–517, 1992.

[68] N. Katz, D.H. Weinberg, and L. Hernquist. Cosmological Simulations with TreeSPH. *ApJS*, 105:19–35, 1996.

[69] N. Katz and S.D.M. White. Hierarchical Galaxy Formation: Overmerging and the Formation of an X-ray Cluster. *ApJ*, 412:455–478, 1993.

[70] G. Kauffmann, J.M. Colberg, A. Diaferio, and S.D.M. White. Clustering of Galaxies in a Hierarchical Universe: II. Evolution to High Redshift. *MNRAS*, 307-3:529–536, 1999.

[71] G. Kauffmann, S.D.M. White, and B. Guiderdoni. The Formation and Evolution of Galaxies within Merging Dark Matter Haloes. *MNRAS*, 264:201–218, 1993.

[72] S.T. Kay, F.R. Pearce, A.Jenkins, C.S. Frenk, S.D.M. White, P.A. Thomas, and H.M.P. Couchman. Parameter tests within cosmological simulations of galaxy formation. *astro-ph*, 9908107:1–22, 1999.

[73] D. Kirkman, D. Tytler, S. Burles, D. Lubin, and J. M. O'Meara. QSO 0130-4021: A Third QSO Showing a Low Deuterium-to-Hydrogen Abundance Ratio. *ApJ*, 529:655–660, 2000.

[74] A.S. Kulessa and D. Lynden-Bell. The Mass of the Milky Way Galaxy. *MNRAS*, 255:105–118, 1992.

[75] C. Lacey and S. Cole. Merger Rates in Hierarchical Models of Galaxy Formation. *MNRAS*, 262:627–649, 1993.

[76] J.C. Lombardi, A. Sills, F. Rasio, and S.C. Shapiro. Tests of Spurious Transport in Smoothed Particle Hydrodynamics. *astro-ph*, 9807290:1–49, 1998.

[77] L.B. Lucy. A Numerical Approach to the testing of the fission hypothesis. *AJ*, 82-12:1013–1024, 1977.

[78] P.J. Maguire. *Astrophysical Gas Dynamics of Structure Formation.* PhD Thesis, Department of Physics, University of Wollongong, Wollongong, Australia, 1996.

[79] J. Makino. Comparison of Two Different Tree Algorithms. *J Comp Phys*, 88:393–408, 1990.

[80] R.O. Marzke, M.J. Geller, J.P. Huchra, and H.G. Corwin. The Luminosity Function for Different Morphological Types in the CfA Redshift Survey. *AJ*, 108-2:437–445, 1994.

[81] C.A. Metzler and A.E. Evrard. A Simulation of the Intracluster Medium with Feedback from Cluster Galaxies. *ApJ*, 437:564–583, 1994.

[82] J.C. Mihos and L Hernquist. Star-forming Galaxy Models: Blending Star Formation into Treesph. *ApJ*, 437:611–624, 1994.

[83] D. Molteni, D. Ryu, and S.K. Chakrabarti. Numerical Simulations of Standing Shocks in Accretion Flows around Black Holes: A Comparative Study. *ApJ*, 470:460+, 1996.

[84] J.J. Monaghan. Smoothed Particle Hydrodynamics. *ARA&A*, 30:543–547, 1992.

[85] J.J. Monaghan and R.A. Gingold. Shock Simulation by the Particle Method SPH. *J Comp Phys*, 52:374–389, 1983.

[86] J.J. Monaghan and J.C. Lattanzio. A Refined Particle Method for Astrophysical Problems. *A&A*, 149:135–143, 1985.

[87] M. Mori, Y. Yoshii, T. Tsujimoto, and K. Nomoto. The Evolution of Dwarf Galaxies with Star Formation in Outward Propagating Super Shell. *ApJ*, 478:L21–L24, 1997.

[88] J. R. Mould, J. P. Huchra, W. L. Freedman, Jr. Kennicutt, R. C., L. Ferrarese, H. C. Ford, B. K. Gibson, J. A. Graham, S. M. G. Hughes, G. D. Illingworth, D. D. Kelson, L. M. Macri, B. F. Madore, S. Sakai, K. M. Sebo, N. A. Silbermann, and P. B. Stetson. The Hubble Space Telescope Key Project on the Extragalactic Distance Scale. XXVIII. Combining the Constraints on the Hubble Constant. *ApJ*, 529:786–794, 2000.

[89] J.R. Murray. Angular Momentum Transport in Simulations of Accretion Disks. *Computational Astrophysics; 12th Kingston Meeting on Theoretical Astrophysics; proceedings of meeting held in Halifax; Nova Scotia; Canada October 17-19, 1996*, 123:78+, 1996.

[90] J.V. Narlikar and T. Padmanabhan. Inflation for Astronomers. *ARA&A*, 29:325–362, 1991.

[91] J.F. Navarro and W. Benz. Dynamics of Cooling Gas in Galactic Dark Halos. *ApJ*, 380:320–329, 1991.

[92] J.F. Navarro, C.S. Frenk, and S.D.M. White. The Assembly of Galaxies in a Hierarchically Clustering Universe. *MNRAS*, 275:56–66, 1995.

[93] J.F. Navarro, C.S. Frenk, and S.D.M. White. Simulations of X-ray Clusters. *MNRAS*, 275:720–740, 1995.

[94] J.F. Navarro and M. Steinmetz. The Effect of Photoionizing Ultraviolet Background. *ApJ*, 478:13, 1997.

[95] J.F. Navarro and S.D.M. White. Simulations of Dissipative Galaxy Formation in Hierarchically Clustering Universes - I. Test of the Code. *MNRAS*, 265:271–300, 1993.

[96] R.P. Nelson and J.C.B. Papaloizou. Variable Smoothing Lengths and Energy Conservation in Smoothed Particle Hydrodynamics. *MNRAS*, 270:1–29, 1994.

[97] M.L. Norman and G.L. Bryan. Cosmological Adaptive Mesh Refinement. *Numerical Astrophysics: Proceedings of the International Conference on Numerical Astrophysics 1998, Tokyo, Japan - Astrophysics and space Library*, 240:19+, 1999.

[98] P.E.J. Nulsen, X. Barcons, and A.C. Fabian. Damped Lyα Absorbers from Dwarf Galaxy Ejecta. *MNRAS*, 301:168–174, 1998.

[99] P.E.J. Nulsen and A.C. Fabian. The role of cooling flows in galaxy formation. *MNRAS*, 277:561–576, 1995.

[100] P.E.J. Nulsen and A.C. Fabian. Gas Processes During the Formation of Galaxies. *MNRAS*, 291:425–436, 1997.

[101] P.E.J. Nulsen and A.C. Fabian. Fuelling Quasars with Hot Gas. *MNRAS*, 311:346–356, 2000.

[102] J. M. Owen, J. V. Villumsen, P. R. Shapiro, and H. Martel. Adaptive Smoothed Particle Hydrodynamics: Methodology. II. *ApJS*, 116:155+, 1998.

[103] T. Padmanabhan. *Structure Formation in the Universe.* Press Syndicate of the University of Cambridge, Cambridge, 1993.

[104] J.A. Peacock. *Cosmological Physics.* Cambridge University Press, Cambridge, 1998.

[105] F.R. Pearce, C.S. Frenk, A. Jenkins, J.M. Colberg, P.A. Thomas, H.M.P. Couchman, S.D.M. White, G. Efstathiou, J.A. Peacock, and A.H. Nelson. Cosmological Galaxy Formation. *astro-ph*, 9906032:1–4, 1999.

[106] P.J.E. Peebles. *Principles of physical cosmology.* Princeton University Press, Princeton, USA, 1993.

[107] S. Perlmutter, G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent, P. G. Castro, S. Deustua, S. Fabbro, A. Goobar, D. E. Groom, I. M. Hook, A. G. Kim, M. Y. Kim, J. C. Lee, N. J. Nunes, R. Pain, C. R. Pennypacker, R. Quimby, C. Lidman, R. S. Ellis, M. Irwin, R. G. McMahon, P. Ruiz-Lapuente, N. Walton, B. Schaefer, B. J. Boyle, A. V. Filippenko, T. Matheson, A. S. Fruchter, N. Panagia, H. J. M. Newberg, W. J. Couch, and The Supernova Cosmology Project. Measurements of Omega and Lambda from 42 High-Redshift Supernovae. *ApJ*, 517:565–586, 1999.

[108] M. Postman and M.J. Geller. The Morphology-Density Relation: the Group Connection. *ApJ*, 281:95–99, 1984.

[109] W.H. Press and P. Schechter. Formation of Galaxies and Clusters of Galaxies by Self-similar Gravitational Condensation. *ApJ*, 187:425–438, 1974.

[110] M.J. Rees and J.P. Ostriker. Cooling, Dynamics and Fragmentation of Massive Gas Clouds: Clues to the Masses and Radii of Galaxies and Clusters. *MNRAS*, 179:541–559, 1977.

[111] D. Ryu, J.P. Ostriker, H. Kang, and R. Cen. A cosmological hydrodynamic code based on the total variation diminishing scheme. *ApJ*, 414:1–19, 1993.

[112] D. Ryu, E.T Vishniac, and W. Chiang. A model for the distribution of the intergalactic medium. *ApJ*, 354:389–399, 1990.

[113] A. Serna, J.M. Alimi, and J.P. Chieze. Adaptive Smooth Particle Hydrodynamics and Particle-Particle Coupled Codes: Energy and Entropy Conservation. *ApJ*, 461:884–896, 1996.

[114] P.R. Shapiro, H. Martel, J.V. Villumsen, and J.M. Owen. Adaptive Smoothed Particle Hydrodynamics with Application to Cosmology: Methodology. *ApJS*, 103:269–330, 1996.

[115] F.H. Shu. *The Physics of Astrophysics, Vol II: Gas Dynamics.* University Science Books, Sausalito, USA, 1992.

[116] I. Smail, R.S. Ellis, M.J. Fitchett, and A.C. Edge. Gravitational lensing of distant field galaxies by rich clusters - II. Cluster mass distributions. *MNRAS*, 273:277–294, 1995.

[117] S.C. Smith, J.L. Houser, and J.M. Centrella. Simulations of Nonaxisymmetric Instability in a Rotating Star: A Comparison between Eulerian and Smooth Particle Hydrodynamics. *ApJ*, 458:236–256, 1996.

[118] R.S. Somerville, G. Lemson, T.S. Kolatt, and A. Dekel. Evaluating semi-analytic halo merging histories. *MNRAS Submitted, astro-ph*, 9807277:1–10, 1998.

[119] R.S. Somerville, J.R. Primack, and S.M. Faber. The Nature of High-Redshift Galaxies. *Physics Reports*, 307:15–22, 1998.

[120] J. Sommer-Larsen, S. Gelato, and H. Vedel. Formation of Disk Galaxies: Feedback and the Angular Momentum Problem. *ApJ*, 519:501–512, 1999.

[121] L. Spitzer. *Physics of fully ionized gases.* Interscience, New York, 1962.

[122] R.J. Splinter, A.L. Melott, S.F. Shandarin, and Y. Suto. Fundamental Limitations on Cosmological Clustering Simulations due to "N-bodies". *American Astronomical Society Meeting*, 191-86:3+, 1997.

[123] R. Spurzem and S.J. Aarseth. Direct Collison simulation of 100000 particles past core collapse. *MNRAS*, 282:19–39, 1996.

[124] M. Steinmetz. Mergers and the Formation of Disk Galaxies in Hierarchically Clustering Universes. *New light on galaxy evolution. The IAU Symposium 171, held at Max Planck Haus in Heidelberg, Germany from 25th-30th June 1995. Edited by Ralf Bender and Roger L. Davies. Published by Kluwer Academic Publishers*, pages 259+, 1996.

[125] M. Steinmetz and E. Müller. The Formation of Disk Galaxies in a Cosmological Context: Structure and Kinematics. *MNRAS*, 276:549–562, 1995.

[126] M. Steinmetz and S.D.M. White. Two-Body Heating in Numerical Galaxy Formation Experiments. *MNRAS*, 288:545–550, 1997.

[127] J.J. Thacker, E.R. Tittley, F.R. Pearce, H.M.P. Couchman, and P.A. Thomas. Smoothed Particle Hydrodynamics in Cosmology: a Comparative Study of Implementations. *astro-ph*, 9809221:1–30, 1998.

[128] P.A. Thomas and H.M.P. Couchman. Simulating the Formation of a Cluster of Galaxies. *MNRAS*, 257:11–13, 1992.

[129] P.A. Thomas and A.C. Fabian. The formation of dark matter in cooling flows. *MNRAS*, 246:156–162, 1990.

[130] K. Thornton, M. Gaudlitz, H. Janka, and M. Steinmetz. Energy Input and Mass Redistribution by Supernovae in the Interstellar Medium. *ApJ*, 500:95+, 1998.

[131] A.A. Thoul and D.H. Weinberg. Hydrodynamic Simulations of Galaxy Formation. I. Dissipation and the Maximum Mass of Galaxies. *ApJ*, 442:480–491, 1995.

[132] A.A. Thoul and D.H. Weinberg. Hydrodynamic Simulations of Galaxy Formation. II. Photoionization and the Formation of Low-Mass Galaxies. *ApJ*, 465:608–616, 1996.

[133] P.B. Tissera, D.G. Lambas, and M.G. Abadi. Analysis of Galaxy Formation with Hydrodynamics. *MNRAS*, 286:384–392, 1997.

[134] V. Trimble. Existence and Nature of Dark Matter in the Universe. *ARA&A*, 25:425–472, 1987.

[135] P. Valageas and R. Schaeffer. The mass and luminosity functions of galaxies and their evolution. *A&A*, 345:329–362, 1999.

[136] F.C. van den Bosch. Semi-Analytical Models for the Formation of Disk Galaxies: I. Constants from the Tully-Fisher Relation. *ApJ*, 530:177–192, 2000.

[137] J.V. Villumsen. A new hierachical particle-mesh code for very large scale cosmological N-body simulations. *ApJS*, 71:407–431, 1989.

[138] K.W. Weiler and R.A. Sramek. Supernovae and Supernova Remnants. *ARA&A*, 26:295–341, 1988.

[139] S.D.M. White and M.J. Rees. Core condensation in heavy halos: A two-stage theory for Galaxy Formation and Clustering. *MNRAS*, 183:341–358, 1978.

[140] S.E. Woosley and T.A. Weaver. The Physics of Supernova Explosions. *ARA&A*, 24:205–253, 1986.

[141] K.K.S. Wu, A.C. Fabian, and P.E.J. Nulsen. The Effect of Supernova Heating on Cluster Properties and Constraints on Galaxy Formation Models. *MNRAS*, 301:L20–L24, 1998.

[142] K. Yoshikawa, Y.P. Jing, and Y. Suto. Cosmological SPH Simulations with Four Million Particals: Statistical Properties of X-ray Clusters in a Low-density Universe. *astro-ph*, 0001076:1–17, 2000.